ADEQUACY OF OFFICER FITNESS REPORTS

BY

JOSEPH F. FOLEY

Thesis
F54

U. S. Naval Postgraduate School
Annapolis, Md.

THE ADEQUACY OF OFFICER FITNESS REPORTS

A study toward improving the
present methods of evaluation
of officer performance in the
Navy

A Thesis

Presented in Partial Fulfillment of the Requirements

for the Degree Master of Science

Library
U. S. Naval Postgraduate School
Annapolis, Md.

By

JOSEPH F. FOLEY, B.Sc.

The Ohio State University

1950

Approved by:

_____
Adviser

Thesis
F54

Library
U. S. Naval Postgraduate School
Annapolis, Md.

12839

TABLE OF CONTENTS

CHAPTER PAGE

PART I

1. THE PROBLEM AND ITS IMPORTANCE 1

NEED FOR A FORMAL METHOD 1

OBJECTIVES OF RATING 2

THE VALUE OF MERIT RATING TO THE NAVY. 4

2. PLAN FOR TREATMENT OF THE PROBLEM 7

PART II

1. SOME PSYCHOLOGICAL FACTORS BEARING ON THE PROBLEM. OF MERIT RATING 9

NATURE OF RATING. 9

SOME DETERMINANTS OF PERCEPTION. 10

FORMATION OF BELIEFS AND ATTITUDES. 12

RETENTION OF BELIEFS AND ATTITUDES. 12

RESISTANCE TO CHANGE OF BELIEFS AND ATTITUDES . . 13

OTHER FACTORS INFLUENCING OBJECTIVITY. 14

Effect of Length of Acquaintance. 14

Relative Difficulty of Judging Different Traits. 15

Generosity Error 15

Error of Central Tendency 16

Logical Error 16

Halo 17

Differences in Purpose of Raters. 18

Differences in Standards 19

SUMMARY. 20

TABLE OF CONTENTS

CHAPTER PAGE

PART I

1. [illegible] 1
[illegible] 1
[illegible] 2
[illegible] 4
2. [illegible] 7

PART II

1. [illegible] 9
[illegible] 9
[illegible] 10
[illegible] 12
[illegible] 12
[illegible] 13
[illegible] 14
[illegible] 14
[illegible] 15
[illegible] 16
[illegible] 17
[illegible] 18
[illegible] 19
[illegible] 20

CHAPTER PAGE

PART II

2. DEVICES AND PROCEDURES IN SERVICE RATING 22

GRAPHIC RATING SCALE 22

Scoring the Graphic Scale. 24

The profile method 24

The direct calculation method. 25

The graphic-computation method 26

The machine method 26

Types of Scores 27

Weighting of Traits. 28

Need for Control of Weighting 30

Advantages of Graphic Scale 31

Disadvantages of Graphic Scale 31

Failure to obtain spread of scores 31

The problem of reliability. 33

The problem of validity. 35

FORCED DISTRIBUTION RATING 36

RANK ORDER METHOD 38

PAIRED COMPARISONS. 38

MAN TO MAN SCALE 40

CHECK LIST 41

COMBINATIONS. 43

POOLING AND REVIEW OF RATINGS 44

Plural ratings. 44

Review of ratings. 46

CHAPTER PAGE

PART II

8. PRINCIPLES [illegible] DRAWING 22
GRAPHIC ARTS [illegible] 22
Scale, the Graphic Scale. 24
The profile method 24
The direct calculation method. 26
The pre-line-computation method 28
The [illegible] method . . . . . [illegible]
Types of [illegible] 29
Signs of [illegible]. 29
[illegible] 30
A [illegible] of Graphic Scale 31
[illegible] Graphic Scale 31
Making [illegible] 31
The [illegible] 33
[illegible] 33
[illegible] 34
[illegible] 34
[illegible] 35
[illegible] 35
[illegible] . . . . . [illegible]
[illegible] . . . . . [illegible]

CHAPTER PAGE

PART II
(Continued)

NEW APPROACH. 47

NUMBER OF TRAITS 48

Mathematical considerations. 48

In favor of a reduced number of traits . . . 49

The case for a larger number 50

Practical considerations. 52

ADMINISTRATION OF RATING PROGRAM 52

Showing ratings to ratees 52

Who should rate. 53

How often to rate 56

Rating the rater 57

TRAINING OF RATERS. 57

Effect of training measured. 58

Effect of certain rater qualities. 59

Substance of Rater Training Program 60

The purpose of rating. 60

Theory of normal distribution 60

Psychological influences. 61

Meaning of terms 61

Practice ratings 61

SUMMARY OF ART IN RATING [illegible] . 62

Principles Applicable to Devices 62

Bias and Objectivity 62

Halo 63

CHAPTER PAGE

PART II
(Continued)

Normal distribution 63

Weighting of traits 64

Display of qualities 64

Combinations. 65

Traits to be excluded. 65

Administrability 66

Principles Applicable to Procedure. 66

Conference with rates. 66

Who should rate. 66

How often to rate 66

Review of ratings 67

Training of raters. 67

Rating raters 67

3. CRITERIA. 68

Importance of Criteria. 68

Inadequacy of Available Criteria 68

Criteria of reliability 68

Typical coefficients of reliability 69

Criteria of validity 70

Typical criteria of validity 72

Ratings by peers as criteria 73

Ratings by peers instead of superiors . . . 75

Future possibilities 76

TABLE II
(Continued)

Normal distribution 53
Adjusting of tables 54
Giving of qualities 54
Conclusions. 55
Tests to be analyzed. 56
Analytic validity 56
Principles applicable to research. 56
Conclusions - for tests. 56
The social tests. 56
How often is test 57
Review of testing. 57
Training of raters. 57
Rater errors 57
3. [illegible] 58
Importance of criteria. 58
Increment of a clinical criteria 59
(Criteria of validity) 59
Typical correlations of validities 59
Length of [illegible] 59
(...) criterion ... criteria) 60
[illegible] 60
[illegible] 61
Study [illegible] 61

CHAPTER PAGE

PART II
(Continued)

4. SERVICE RATING IN THE ARMED FORCES 80

Some Early Navy Forms 80

Subsequent developments. 80

The Present Navy Form 81

Physical makeup 83

Use of form. 83

Some analyses of Navy ratings. 84

Evaluation 87

The Army Rating Form. 89

Physical makeup and use. 91

The forced choice theory 92

Some analyses of Army ratings. 93

Evaluation 94

The Air Force Report. 99

Description and use of booklet 99

The rating sheet 100

Evaluation 101

The Marine Corps Rating Form 104

Analysis. 106

Evaluation 106

SUMMARY. 108

PART II
(Continued)

4. [illegible] 3

Some Early Ewe 4

Homogeneity in Anlo 4

The Ewe Speaking Area 5

Physical Design 6

The of Ewe 8

Some Problems in Ewe Religion 8

Religion 87

The Ewe Godhead 94

Ewe and Nature Gods 91

The Lesser Godhead Ahoyo 96

The Ancestors as Ewe Will 99

Religion 14

The Ewe Force Report 99

Reincarnation and Ewe Ancestors 11

The Medicine Men 104

Religion 1

Ewe Large Medicine Men 1

Sorcery 1

Religion 1

CHAPTER PAGE

PART III

1. CONCLUSIONS. 109

2. RECOMMENDATIONS 112

First Recommendation. 112

Form 113

Second Recommendation 114

Third Recommendation. 114

Fourth Recommendation 114

Subsidiary Recommendations. 115

Information to be Elicited. 116

Significant areas of information. 117

Specificity of Rating Device 118

Information to be Excluded. 120

Length and Complexity 120

Machine Processing 122

REFERENCES 124

BIBLIOGRAPHY 137

APPENDIX A. SAMPLES OF TYPES OF RATING FORMS. . . 143

CONTENTS

CHAPTER III

1. [illegible] 109

2. RECOMMENDATIONS 118
First Recommendation 118
Form 123
Second Recommendation 124
Third Recommendation 124
Fourth Recommendation 124
Delivery Sect. 40 Bylaws 126
Information on Ex-Officio 126
Significant area of [illegible] 127
Possibility of Radio Series 128
Information to [illegible] 130
Length and Complexity 130
Mobile Processing 132
[illegible] 134
Bibliography 135
[illegible] 141

CHAPTER PAGE

PART III
(Continued)

APPENDIX B. AN EARLY NAVY RATING FORM 147

APPENDIX C. A MORE RECENT NAVY RATING FORM 149

APPENDIX D. PRESENT NAVY RATING FORM 151

APPENDIX E. ARMY RATING FORM 153

APPENDIX F. AIR FORCE RATING WORK BOOKLET
(SAMPLE SHEET) 156

APPENDIX G. AIR FORCE RATING SHEET 158

APPENDIX H. MARINE CORPS RATING FORM 160

APPENDIX I. GLOSSARY OF TERMS 162

CHAPTER PAGE

PART III
(Continued)

APPENDIX B. AN EARLY NAVY RATING FORM 147

APPENDIX C. A MODE -RONEY NAVY RATING FORM 149

APPENDIX D. PRESENT NAVY RATING FORM 151

APPENDIX E. ARMY RATING FORM 155

APPENDIX F. AIR FORCE RATING FORM SCALES (SAMPLE SHEET) 156

APPENDIX G. AIR FORCE RATING SHEET 158

APPENDIX H. MARINE CORPS RATING FORM 160

APPENDIX I. GLOSSARY OF TERMS 164

LIST OF FIGURES

FIGURE PAGE

1. A DIAGRAM SHOWING HOW A PRESCRIBED DISTRIBUTION APPROXIMATES NORMAL DISTRIBUTION 37
2. FORM TO AID RATER IN PREPARING FOR POST RATING INTERVIEW 54
3. PROCEDURE FOR POST RATING INTERVIEW 55
4. GRAPHIC PRESENTATION OF DISTRIBUTION DATA OBTAINED FROM STUDY OF NAVY OFFICER RATINGS . 86
5. DISTRIBUTION OF OVERALL RAW SCORES ON ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1, ALL GRADES 96
6. DISTRIBUTION OF OVERALL RAW SCORES ON ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1, FIELD GRADE 97
7. DISTRIBUTION OF OVERALL RAW SCORES ON ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1, COMPANY GRADE 98

LIST OF FIGURES

FIGURE		PAGE
1.	A DIAGRAM SHOWING HOW A PERFORMANCE DISTRIBUTION APPROXIMATES NORMAL DISTRIBUTION	37
2.	FORM TO AID SUTER IN PREPARING FOR POST RATING INTERVIEW	84
3.	PROCEDURE FOR POST RATING INTERVIEW	85
4.	GRAPHIC PRESENTATION OF DISTRIBUTION DATA OBTAINED FROM STUDY OF NAVY OFFICER RATINGS .	86
5.	DISTRIBUTION OF OVERALL RAS SCORES IN ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1, ALL GRADES	95
6.	DISTRIBUTION OF OVERALL RAS SCORES IN ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1, FIELD GRADE	97
7.	DISTRIBUTION OF OVERALL RAS SCORES IN ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1, COMPANY GRADE	98

PART I

INTRODUCTION

Chapter 1

THE PROBLEM AND ITS IMPORTANCE

The problem of rating or judging one's subordinates is as old as Organization itself. Many writers have stressed the point that judging the employees under him is an integral part of the job of the supervisor. May[1] argued that consciously or not, the supervisor rates his men whenever called upon to choose between them for special assignments, special awards, promotions, or reductions in force. Jucius[2] considers that the rating of employees is one of the oldest and most universal practices of management and that there never was a time when supervisors and leaders did not judge the relative value of employees. Tiffin[3] emphasized that though the term "merit rating" (service rating) is of recent origin, the judging of employees by their supervisors is not.

NEED FOR FORMAL METHOD

If many writers have agreed that service rating cannot be dispensed with, an equal number are in consonance as to the need for a formal method of making and recording these judgements. Marble[4] proclaimed that the service rating device protects the employee from "managerial capriciousness" in the determination of his service standing.

White[5] taught that one of the purposes of the formal efficiency rating is "to protect employees against the snap judgements, prejudice, and ill will of some supervisors". Tiffin[6] feels that the emergence of the formal rating system has "involved the transfer of ratings from haphazard, random, and frequently irresponsible judgements of supervisors made perhaps during the heat of a quarrel, to ratings made calmly, deliberately, systematically....". In addition to this need for minimizing the effects on the ratings of the frailties of the rater, Jucius[7] sees further value in the formal recorded rating, in providing a means for preserving a measure of an employee's worth so that even in a large industry the treatment received by all employees and the opportunities opened to them will be consistent throughout the organization.

OBJECTIVES OF RATING

The objectives of rating are manifold. Principally, of course, the purpose is to secure a recorded judgement of the employee's value to the organization. But underlying this and giving it direction and motive, are many specific aims. Mention has been made of the utility of ratings in affording a basis for choice between one's subordinates for special assignments, special awards, promotions, and reductions in force. Less obvious but no less important, is the use suggested by Yoder[8], of ratings as criteria for evalu-

While [illegible] tends to [illegible] the formal [illegible] clause, really is "to protect [illegible] the [illegible] in [illegible] [illegible] [illegible] [illegible] [illegible]". Tiffin tends [illegible] the [illegible] as the formal [illegible] [illegible] "involved [illegible] [illegible] of [illegible] from [illegible] problem, and frequently [illegible] [illegible] [illegible] some [illegible] [illegible] [illegible] [illegible], [illegible] [illegible] made [illegible] [illegible] [illegible] [illegible]...". [illegible] [illegible] to [illegible] [illegible] [illegible] [illegible] the [illegible] [illegible] [illegible] [illegible]

ating the effectiveness of other personnel practices. Tiffin[9], Pigors and Myers[10], and Mosher and Kingsley[11] made much the same point when they listed ratings as a source of information on which to base programs for employee improvement, training, and development. Associated with this is the use of the data from ratings as a foundation and guide for studies to improve screening and selection procedures. Further, operation of a good rating plan results in improvement of the supervisor, as mentioned by both Tiffin[12] and Jucius[13], and therefore such improvement may be considered as one of the purposes of the program.

Not specifically mentioned in the literature is the use of the service rating as an instrument to aid in the fostering of good discipline and morale. Much good work and many minor deficiencies are not deserving of special action other than as reflected in the marking of the service rating blank. Yet good work must be rewarded and brought to the attention of the employee and management, and the service rating provides a means of doing this. Though the effect on the employee may be vitiated somewhat by the fact that the psychologically sound principle of the "immediacy of the reward" is violated, the dependance on the service rating as a form of reward is justified on pragmatic grounds.

[illegible] the effectiveness of other personnel practices. Tiffin[9], Flores and [illegible][10], and Rogers and Alexander[11] made much the same point [illegible] as a source of information on which to base decisions for employee improvement, training, [illegible] promotion. Reports on also tend to the use of the [illegible] as a formation and guide for action in improving [illegible] procedures. Further, [illegible] the results [illegible] in improvement of the supervisor, as [illegible] by both Tiffin[12] and Jenkins[13], and therefore such improvement may be considered as one of the purposes of the program.

The specificity mentioned in the literature is the use of the service rating as an instrument in aid in the learning of job principles and skills. Good work and many other definitions [illegible] of special [illegible] rating plans. You [illegible] and [illegible] to the [illegible] of the service [illegible] effect [illegible] Logan [illegible] [illegible] [illegible]

THE VALUE OF SERVICE RATING TO THE NAVY

The considerations which serve to render service ratings so valuable to large industries are effective to an even greater degree in the personnel administration of the Naval Service. Three conditions operate to make this true. First, there is the fact of sheer numbers. The officer corps of the Regular Navy is composed of 43,939 individuals[14] while the officers of the Naval Reserve total up to a much larger figure. Second, the problem posed by size is complicated by the twin factors of wide dispersion over large areas of the earth, and mobility within and beyond those areas. In fact, it is safe to say that Naval units are likely to be found anywhere except possibly behind the "Iron Curtain", the currently popular term for the boundary between Russian dominated territory and the rest of the world. Finally, the difficulties of Naval authorities are not limited to the problem of size and ubiquity. An even more severe strain is placed on administration during periods of rapid expansion in times of National Emergency and during the process of demobilization when the emergency has passed. With final authority for all administrative action necessarily centralized in the Navy Department in Washington, and with so many individuals involved in numerous widely separated and mobile groups, personal knowledge on the part of those responsible for admin-

[illegible]

[illegible]

The counterpoising [illegible] serve to render service [illegible] as valuable to large [illegible] are effective as a [illegible] even greater [illegible] in the permanent [illegible] of the Naval service. These conditions [illegible] to make this [illegible] Pilot, [illegible] to be [illegible] at sea [illegible]. The officer corps of the regular Navy is composed of 4,000 individuals[illegible] while the officers of the Naval reserve total [illegible] to a much larger figure. [illegible] conditions [illegible] by also in conjunction [illegible] the [illegible] of the [illegible] upon large groups of the [illegible], as [illegible] and [illegible] those areas. In fact, it is [illegible] that Naval units are lined [illegible] "Blue Corps", the [illegible]. [illegible] the boundary between [illegible] and the rest of the [illegible]. Finally, the [illegible] of Naval [illegible] this [illegible] to [illegible] in time [illegible] life. [illegible] during [illegible] of a [illegible] [illegible] and [illegible] of [illegible] the [illegible] and [illegible]. [illegible] all [illegible] [illegible]

istrative action, of the qualities of even a small number of the officers of the Navy is impossible.

The Navy Department recognizes the importance of service ratings, or fitness reports. Ample evidence of this is contained in the following passage from Navy Regulations.[15]

> Article 1701. Fitness Reports and Records of Officers.
>
> 1. The fitness of an officer for the service, with respect to promotion and assignment to duty, is determined by his record, Reports of fitness are decisive in the service career of the individual officer, and have an important influence on the efficiency of the entire service.....

The Navy's "OFFICER'S FITNESS REPORT", Form NAVPERS 310 A (Rev 6-45), is intended to fill the need for a merit rating in the Service, but unfortunately it is a partial failure. First of all, it is subject to significant observer errors, like most rating devices and, disappointingly, its design omits several features which authorities consider will eliminate or reduce such errors. In addition, its makeup does not provide for the use of the latest techniques in machine scoring, sorting, and tabulation. This is an important weakness in view of the growing importance of maximum efficiency implicit in the concept of readiness for "Total War". Although small numbers of the fitness reports are not laborious to prepare, any considerable study of even a moderate number of them, such as must be performed when making selections for promotion or special assignments, or

[illegible]

[illegible]

[illegible]

for evaluating other personnel procedures, is inefficient and time consuming under present conditions.

for evaluating other personnel procedures, is inefficient and time consuming under present conditions.

PART I

Chapter 2

PLAN FOR TREATMENT OF THE PROBLEM

The design and use of service ratings is a task in psychological measurement and the designer's most formidable obstacle has been the difficulty of eliminating or controlling the "human error" or bias of the rater. This paper will, therefore, examine rather thoroughly the psychological factors bearing on the question in an effort to show the nature and extent of this "human error" and perhaps assist the reader in gaining insight into the problem.

Much valuable research has been and is being performed in the field of service rating design and use, and a great deal can be learned from a review of the reported results. Consequently, a chapter will be devoted to the description and discussion of the more important types of rating devices and rating procedures in use. This discussion will include weighing of the relative advantages and disadvantages of the several devices and systems and consideration of the various error-producing psychological factors which the measures described are designed to combat. From this treatment a body of "best practices" will be assembled.

The four services which are included under the general name of "Armed Forces", each use a different type of officer's service rating and each device has its good and

[illegible]

[illegible]

[illegible]

The [illegible] of [illegible] is [illegible] psychological [illegible] obstacle [illegible] controlling the "[illegible]" [illegible] water. This paper will therefore [illegible] the psychological factors [illegible] the [illegible] to allow us to the [illegible] and [illegible] "[illegible]" and [illegible] model the [illegible] problems.

[illegible] theory [illegible] and is only particularly [illegible] [illegible] [illegible] problem [illegible] [illegible] to [illegible] [illegible] [illegible] use. This [illegible] [illegible] [illegible]

its bad points. These military service ratings will be described, analyzed, and judged, in the light of the aforementioned "best practices" and, based on this assessment, as well as on the "best practices", a plan for a new improved officer's fitness report form will be proposed for adoption by the Navy. No actual work of construction of the device will be attempted, as such is beyond the scope of this thesis.

As has been stated, much research has been done in this field. In the review of the reported results of that research, credit will be given to the scientists whose material is quoted or referred to wherever possible. It is desired, in addition, to make specific mention here of the works of Grant Sheridan Baze and of Wayne Gilmer Samples. In 1947, Mr. Baze submitted a master's thesis to the Graduate School of The Ohio State University entitled A Rating Scale for Marine Non-Commissioned Officers of the First Three Pay Grades. A year later, Mr Samples submitted a master's thesis to the same authority entitled A Study of Merit Rating Methods for Professional Employees. Each writer developed a rating scale which he recommended for the purpose stated in the title of his thesis. Although the general type of scale produced by both Mr. Baze and Mr. Samples (the graphic rating scale) could be used as a Naval officer's efficiency rating, the particular scales developed and reported would not be suitable and therefore will not be discussed further.

PART II

Chapter 1

SOME PSYCHOLOGICAL FACTORS BEARING ON THE PROBLEM OF SERVICE RATING

NATURE OF RATING

Heretofore in this thesis, service rating has been described as the judging of the ratee's worth to his organization. Reliance is placed on judgements in this important field of measurement because of lack of better method. To paraphase Yoder[1] it may be said that many human qualities such as leadership, initiative, persistence, self-control, and others are important to vocational success but objective measures of such characteristics do not exist.

These judgements are reached as a result of the rater's observation or perception of the typical performance, or the characteristic actions or behaviors of the ratee. It has been said that, "Nowhere so much as in judging people is perception likely to be so biased."[2] Personality is generally thought of by scientists as the total pattern of an individual's reaction to his environment. In truth Stagner[3] has defined personality as the integral of an individual's characteristic responses. It is evident, therefore, that the service rating, in assessing the typical performance of the ratee, is attempting to measure an aspect of his personality.

A moment's reflection will bring the realization that

PART II

Chapter I

SOME PSYCHOLOGICAL FACTORS BEARING ON THE PROBLEM OF ADJUSTING NATURE

NATURE OF A THING

Heretofore in this thesis, service work has been described as the "giving of the water" work to his organization. Alliance is placed on [illegible] in this important field of management a sense of lack of better method. The perspicacious Yoder[1] is [illegible] such very human qualities such as leadership, initiative, personnel, self-control, and others are important to [illegible] successful subjective measures of such characteristics is not crucial.

These intangibles are [illegible] as a result of [illegible] water's observation of personnel [illegible] the [illegible] performance, or the characteristic action or behaviors of the water. It has been said, that, "Water [illegible] or in [illegible] people is perceptible likely to be [illegible] [illegible] generally thought of as [illegible] [illegible] individuality [illegible] to [illegible] environment. In [illegible] has defined personality as [illegible] [illegible] [illegible]

personality is a complex entity; just how complex is revealed in the findings of Allport and Odbert[4] that there are 17,953 words in the English language used to describe the behavior of one person as distinguished from another. Even when allowance is made for overlap in meaning among these words, the range of differences in behavior provided for remains very great indeed.

Due to the interrelation between a man's perceptions and his beliefs and attitudes, which will be discussed more fully below, the service rating is more than a record of personality as perceived. Under most systems, inextricably included in his concept of the ratee is a description of the rater's beliefs and attitudes toward the ratee.

The value of a formal rating system has been discussed. Such a system will require the rater to describe his concept of each ratee on a separate form. A well designed form will provide guidance and assistance to the rater in his striving for objectivity as well as a place to record his judgements. Fuller appreciation of the acuteness of the need for this guidance and assistance will come as we discuss on succeeding pages some of the more important psychological influences operating to impair the objectivity of the rater's judgements.

SOME EFFECTS OF PERCEPTION

Social psychologists teach that a person's perception of what takes place around him is profoundly affected

[illegible]

by factors within him and that many of these interior influences are themselves the product of the individual's prior perception and learning. An example of this theory as well as a clue to the nature of some of the interior modifiers may be found in the following quotations:

> How we perceive the world is a product of memory, imagination, hearsay, and fantasy as well as what we are actually "perceiving" through our senses.[5]

> Data are perceived and interpreted in terms of the individual perceiver's own needs, own emotions, own personality own previously formed cognitive patterns.[6]

The reader will readily see the truth in the foregoing and can perhaps recall actual instances in which portions of it have been borne out by experience. In addition there has been much experimental proof. However, only a few of the more important experiments can be cited here because of space limitations.

It was mentioned that imagination, emotions, and "previously formed cognitive patterns" (beliefs and attitudes) affect perception. Murray[7] found this borne out in his experiment in which a group of girls perceived maliciousness to a greater extent in a man's pictured face after they had played a game of "murder". Leuba and Lucas[8] found similarly that a person's mood significantly influenced his perception of moods in the people in pictures shown to him. Another investigator, Heider, reported that a person may be thought guilty of a crime, because he "looks as if he could have

by factors within him and that many of these learning influences are themselves the product of the individual's prior perception and learning. An example of this theory as well as a clue to the nature of some of the learning modifiers may be found in the following quotations:

> Man no perceive the world in a context of memory, imaginative learning, and feeling as well as what we are actually "perceiving" through our senses.[6]
>
> Data are perceived and interpreted in terms of the individual perceiver's own needs, his conditions, his personality and previously learned capacity performance.[7]

The reader will readily see the crux in the terms going on- our perhaps inborn motive- learning in various conditions of it and some other and by experience. In addition there are some such experimental model. However, only a few of the very important experiments which have bearing of space limitations.

A data mentioned that "perceiving, conditions, and "practice" have regarding "learning" (pain-1) and (illusory) effect perception. Thus, learning [illegible] perceiving [illegible] perceived [illegible] to [illegible] place [illegible] [illegible] [illegible]

committed this crime" or because he has the disposition to do such a thing[9]. And finally Zillig[10] demonstrated that people tend to fail to see the mistakes committed by people they like while at the same time attributing mistakes, without justification, to persons they dislike.

FORMATION OF BELIEFS AND ATTITUDES

Another aspect of the relation between perception and beliefs and attitudes is to be found in the following passage from Krech and Crutchfield:[11]

> A belief (and an attitude also)[12] is a pattern of meanings of a thing;...It is a product of original perceptions, of the reorganization those perceptions have undergone, of the cognitions created by such reorganization, etc.

That is to say, not only is perception influenced by beliefs and attitudes but they are in turn dependent on perception for the raw materials from which they are made.

RETENTION OF BELIEFS AND ATTITUDES

What about the effect of the passage of time on beliefs and attitudes? It is generally understood that there will be some forgetting. Krech and Crutchfield[13] admitted that what one knows and believes changes with the passage of time but added the valuable thoughts that the effect of time appears "...not only in errors of omission but frequently in errors of commission as well" and "...facts are created through the process of creative forgetting." Adopting the language of Krech and Crutchfield, one may add to the

committed this crime" or, assuming he has the disposition to do such a thing[9]. Jan Plaskiewicz[10] demonstrated that people tend to fail to see the obvious: committed by people they like while at the same time attributing aberration, without justification, to persons they dislike.

PERCEPTION OF BELIEFS AND ATTITUDES

Another aspect of the relation between perception and beliefs and attitudes is to be found in the following passage from Kroch and Crutchfield:[11]

> A belief (and an attitude also) is a pattern of meaning, of a thing...It is a product of original personalities, of the organizations since childhood, the experiences, of the cognitions arrived at, and the organizations, etc.

There is so, not only in perception influenced by beliefs and attitudes but, in fact, it is more dependent on perception for the psychological process to take place.

PERCEPTION OF BELIEF AND ATTITUDE

What alone the effect of the process of like or dislike are attributed. If I know will motivation, but I know will be more for action. From Jan Crutchfield[12] admitted that what can there be "objects upon we rely the feeling of this sub-, but, psychologists consider the effect of time spent "...but, if expect of emotion in fundamental to the process of emotional-cognitive. "...there is operation based the process of emotional-cognitive. ' ... ing a ... of time ... all fields, the ... in

foregoing the process of "selective forgetting".

Clearcut experimental demonstration of "errors of commission" may be found in Bartlett's[14] experiment in which his subjects, when required to retell an unusual and somewhat ambiguous story, tended to change the unfamiliar content into more familiar material. "Selective forgetting" was recorded by Levine and Murphy[15] when they found that their anti-Communist subjects remembered anti-Communist material better while the pro-Communist subjects retained pro-Communist material more completely. The same study also demonstrated the influence which one's attitudes have on perception, for the subjects learned poorly that material on Communism which was at variance with their views.

RESISTANCE TO CHANGE OF BELIEFS AND ATTITUDES

Though there is a tendency for beliefs and attitudes to change with time, there is a concurrent tendency for them to <u>resist</u> change. This may be thought of as a resistance toward <u>external</u> influences which might cause change. Such a phenomenon is to be expected in the light of the findings of Levine and Murphy already mentioned for there, the information which was contradictory to the views of the subjects was poorly perceived and quickly forgotten. Thus we see that beliefs and attitudes have "self-protective devices"[16] and that they "not only twist and warp data and thus maintain themselves but create new data for self-incorporation and thus grow in intensity."[17]

[illegible] the process of "[illegible]".

[illegible] experiment [illegible] of "[illegible] of consciousness" may be found in [illegible] experiment in which the subjects, who [illegible] to be [illegible] and somewhat analgesic [illegible] to [illegible] the "[illegible]" contact into [illegible] [illegible]. "[illegible] Transparency" was recorded by [illegible] [illegible] when they [illegible] [illegible] [illegible]

OTHER FACTORS INFLUENCING OBJECTIVITY

Effect of Length of Acquaintance

It is obvious that the rater's judgement of the ratee is of little value unless the former individual has been acquainted with the latter one sufficiently long. Yet, as warned by Burtt,[18] there comes a time when, as the length of the acquaintance increases, the accuracy of the ratings decreases. Burtt ascribes this phenomenon to, (1) the natural reluctance of the supervisor to imply that his subordinate has not improved under his leadership, and, (2) the equally natural tendency of the supervisor to identify himself with the older subordinates who are more nearly like him in age and experience. Stockford and Bissell[19] found mathematical proof of these tendencies and commented, "...the longer one has known another, the more acceptable becomes the other's personality."

Another aspect of acquaintance affecting the accuracy of judgements is that determined by the conditions under which the acquaintance was made and maintained. A supervisor's attitude toward a subordinate whom he knows only on the job is likely to be quite different from his feelings toward one with whom he associates while off duty. Experimental support of this generalization may be found in Billig's demonstration, already mentioned on page 15, of how people tend to overlook mistakes made by persons they like.

OTHER FACTORS "MEANINGLESS" BECAUSE OF

Effect of Length of Acquaintance

It is obvious that the exact judgment of acquaintance is of little value unless the factor initial was been acquainted with the latter one sufficiently long. Let, as assumed by Smith, it take place a time when, as the length of the acquaintance increases, the judgment of the relative quantities Brett assigns this increase. So, (1) the original relationship of the acquaintance to Lady [illegible] acquaintance has not improved under his leadership, and, (2) the original motivational tendency of the motivation to [illegible] with the acquaintance or acquaintances has grown hardly like [illegible] perhaps. However, [illegible] of these factors [illegible] known motives, the more [illegible] personality.

[illegible]

Relative Difficulty of Judging Different Traits

It has been found that raters can judge some traits more reliably than others. This phenomenon may be attributed to factors inherent in both the raters and the qualities being considered. The raters, due to their own disparate cultural backgrounds, may be uncertain as to the exact meanings of the terms used on the rating forms, or they may be doubtful as to just what overt behavior indicates the presence of a given trait. When considering the traits, one finds that those which are simple and less complex, and those which cause behavior culminating in tangible results, are more accurately estimated than are the complex characteristics, or those which leave little concrete evidence of their existence. Stockford and Bissell reported, for instance, that "dependability" and "quality of work" were more reliably judged than "morale" and "cooperation".[20] These results support earlier similar findings of Hollingworth.[21] A further difficulty attributable to the traits themselves is touched upon by Stagner in the words, "Things which have the same name are not necessarily the same; e.g., courage in facing a wild animal is not the same as courage in facing adverse public opinion."[22]

Generosity Error

When it comes to describing their fellow men, human beings are essentially kind. Wisson[23] found this in connec-

Relative difficulty of
Indian Literary Tasks

[illegible]

tion with the research incident to the design of the present rating system of the U. S. Army. His resulting comment was, "Though not a necessity of the logic involved, those items which tend to be used most often, i.e., are generally 'preferred' by raters in describing others, are invariably more favorable items--nice things to say." Cronbach named this tendency "generosity error" and added, "It is common to find 60 to 80 percent of an unselected group rated 'above average' because of the urge to speak favorably if possible.[24]

Error of Central Tendency

Related to generosity error is the tendency of the rater to mark a ratee about whom he has insufficient information "average" or near "average" rather than confess his lack of knowledge. Guilford[25] named this phenomenon the "error of central tendency". Unquestionably, when forced to guess as to the qualifications of an individual, one will enjoy the greatest probability of being correct if he guesses that the ratee is average. But service ratings should be based on observation and judgement and guessing is out of place.

Logical Error

Many raters will reason that because the ratee exhibits a certain trait to some degree, he also possess certain other traits to a similar degree. To their way of thinking, all of that particular group of traits are really the same thing or at any rate, they constitute a syndrome.

Such thinking is fallacious and we are indebted to Greene for focusing attention on it and aptly naming it "logical error".[26]

There was some indication in the Stockford and Bissell study previously mentioned, that the relative proxi ity of certain traits on the scale will have an influence similar in effect to logical error. In other words, the judgement made on a trait affected slightly but measurably the judgements made on adjacent traits. The authors stated that "....the differences in the inter-correlations between ratings of the same traits varied from +.01 to +.26 depending directly upon the change in proximity of the traits from one scale to another....."[27]

Halo

Closely related to logical error in its cause is the "halo" effect. It leads the rater to assume that because his general impression of the ratee is good (or bad), the ratee's showing in all, or in a majority, of the traits on the rating scale is good (or bad). Halo is generally deplored by authorities in rating, with one exception, and, as will be discussed in the next chapter, much design effort has been expended on the problem of how to prevent or minimize it. The exception referred to above is Walter V. Bingham who feels that, to some extent, a halo effect is acceptable and proper, indeed inevitable. In support of this belief he wrote,[28]

such a finding is fallacious and is an indication of [illegible] for focusing attention on it and making it "[illegible] error".[illegible]

There was some indication that [illegible] and [illegible] study probably [illegible] that the [illegible] of certain traits on the [illegible] will be a [illegible] in effort to logical error. [illegible] judgments made on a trait affected actual [illegible] judgments made on adjacent traits. [illegible] found that "..... a difference in [illegible] between ratings of the same trait varied from +.22 to +.54 depending [illegible] upon the place in prominence of an trait from one [illegible] to another....."[illegible]

Halo

Obviously related to logical error [illegible] "halo" effect. It [illegible] general [illegible] above [illegible] [illegible] [illegible]

> ....it is not the rater alone whose reactions to the candidate are in question. He is but typical of others.....who will react to the subject, not as a bundle of isolated traits, but as a person with certain duties. The judgements and responses of all these people will unconsciously and inevitably manifest a halo effect, which is, in part at least, valid.

Differences in Purposes of Raters

It is generally agreed that praise is more effective than reproof in most situations. Definitely greater motivation resulted from praise among school children in an experiment reported by Hurlock.[29] Sophisticated raters are aware of this phenomenon and, unless they guard against it, they may allow this knowledge to influence their ratings to the extent that they give higher marks than are deserved on the basis of performance. Some may even go so far as deliberately to give higher markings in the expectation that such marks will spur the ratee to greater effort. The rater would describe this practice as "giving the ratee something to live up to" and would seek to justify it on the ground that he was improving performance by boosting the morale of the ratee.

The effects of frustration have been demonstrated experimentally and well reported. The individual subjected to frustrating circumstances has the choice of resorting to either adaptive or maladaptive behavior. Desirable forms of adaptive behavior which are most likely to be encountered in the rating situation include such effects as intensification of

effort and reorganization of one's perception of the problem. Maladaptive behavior in the rating situation probably would include such adjustments as aggression, regression, withdrawal, and rationalization.

Many raters, even those lacking in formal psychological education, are cognizant of the possible results of frustration, at least to the extent that they recognize that, though frustration may cause increased effort, it may cause instead, withdrawal or quitting. Realization of this may lead a rater to leniency in rating because of a desire, or preference, for relying on the known good effects of praise rather than risking the induction of the maladaptive response of withdrawal by imposing the frustrating experience of a low rating.

Differences in Standards

It has been said that "Individuals differ with respect to almost any human attribute we are to measure."[30] Private standards of comparison are a personal attribute, a component of personality, and, therefore, will be found to differ from one individual to the next. Like other personality factors, standards of comparison are a product of the individual's innate characteristics and all the countless influences of his environment. The importance here of differences in private standards of comparison lies in their influence on the rater's judgements. Variations in the standard of comparison obviously will be reflected in variations in measurements.

effort and reorganization of one's perception. If the problem of maladaptive behavior in the testing situation probably would include such adjustments as rationalization, regression, withdrawal, and rationalization.

Many raters, even those looking in formal psychological education, are cognizant of the possible results of frustration, at least to the extent that they recognize that, though frustration may cause increased effort, it may cause instead, withdrawal or quitting. Realization of this may lead a rater to hesitancy in rating because of a desire, or preference, for relying on the more good effects of praise rather than risking the induction of the maladaptive response of withdrawal by imposing the frustrating experience of a low rating.

Differences in Training

It has been said that individual differences respect to almost any human [illegible] [illegible] measure. Not [illegible] standards of comparison are a personal attribute, a component of personality, and, therefore, will be [illegible] to [illegible] one individual to the next. Like other personality [illegible], standards of comparison are a function of the individual's innate capabilities and [illegible] [illegible] of his surroundings. [illegible] [illegible] [illegible]

SUMMARY

We have seen that the service rating process is an attempt to measure an aspect of personality. It does this by having the rater endeavor to describe and record his perception of the ratee. The point was made that perception is likely to be particularly biased when judging people. Inducing this bias are such powerful psychological influences as the perceiver's memory, imagination, emotions, beliefs, and attitudes. The interrelation between these factors and perception was discussed with emphasis on the process in which, (1) beliefs and attitudes, formed, as they are, from biased perception, are retained and further biased by selective and creative forgetting and, (2) subsequent perception is rendered further biased and selective with the result that beliefs and attitudes tend to grow in strength. Further discussion described how inaccuracy in service ratings can come from long or friendly acquaintance between rater and ratee, variations in the difficulty of judging different traits, the innate desire of most men to speak well of others, differences among raters in the results they hope to achieve with ratings, and differences among raters in standards of comparisons.

This is the material and these are the conditions with which the designer and constructor of service rating devices and procedures must work. He is attempting to measure that which is difficult to define and which, in the

SUMMARY

We have seen that the service rating process is an attempt to measure an aspect of personality. It does this by having the rater endeavor to describe the ratee's perception of the rates. The point was made that perception is likely to be particularly biased when judging people. Among such biases are such general psychological influences as the perceiver's mood, intention, emotions, beliefs, and attitude. The interpretation between these factors and perception was illustrated with emphasis on the process in which, (1) beliefs and attitudes, formed, as they are, from biased perceptions, are hardened and further biased by selection and cognitive formatting and, (2) subsequent perception is motivated further biased and slightly distorted as the result that beliefs and attitudes tend to grow in strength. Another discussion dealt with the [illegible] problem arising from low or [illegible] rater-ratee rating, [illegible] in the [illegible] of judging different traits, the halo [illegible] of [illegible] the [illegible] of [illegible] [illegible] rater [illegible] [illegible] [illegible]

present stage of development of the science, can be measured in no other way. He is using as agents, creatures whose performance is unreliable and whose purposes in rating often differ. And finally, he is measuring against differing standards.

present stage of development of the science, can be measured in no other way. As is seen as sports, crescendos whose performance is unreliable and whose purposes in telling often differ. And finally, by is measuring against different standards.

PART II

Chapter 2

DEVICES AND PROCEDURES IN SERVICE RATING

In the quest for a solution to the difficult measurement problem discussed in the preceding chapter, many different rating forms and procedures have been tried out. However, they are all similar in that each is essentially a means of obtaining a written description of the ratee in terms of traits or behaviors presumed to be indicative of the presence of those traits.

GRAPHIC RATING SCALE

The most widely used rating form is the graphic rating scale. In its simplest form as shown in appendix A, it consists of a list of traits with short lines or graduated scales opposite each trait. One end of each trait scale represents possession of that trait to a maximum degree while the opposite end corresponds to a minimum degree of possession. The form is used in rating by "checking", or otherwise marking, a point somewhere along each trait scale which is judged to indicate the degree to which the ratee possesses the corresponding trait. This builds up a description of the ratee in terms of the traits marked.

As stated in the foregoing, the ends of the trait scales represent maximum and minimum possession respectively of the indicated traits. But most graphic rating forms go

PART II

Chapter 2

DEVICES AND TECHNIQUES IN [illegible] ACTING

In the quest for a solution to the difficult measurement problem discussed in the preceding chapter, many different rating formulas, recommendations, and trial [illegible]. However, they are all similar in that each is essentially a means of obtaining a written numerical [illegible] of the rates or terms of traits or behaviors presumed to be indicative of the presence of some trait.

GRAPHIC RATING SCALE

The most widely used rating form is the graphic rating scale. In its simplest form, as shown in Exhibit A, it consists of a list of traits with short lines or graduated scales opposite each trait. One end of each trait scale represents possession of that trait to a maximum degree while the opposite end corresponds to a minimum degree of possession. The form is used in rating by "checking", or marking a point somewhere along each trait scale which is judged to indicate the degree to which the rated behavior [illegible] the appropriate trait. This marking is usually done [illegible] in terms of the traits [illegible].

As stated in [illegible], the [illegible] of the trait scales [illegible] [illegible] [illegible]

go further than this. It is common practice to provide guidance for the rater in his choice of the spot to be checked on each trait scale. Varying degrees of possession of the traits are indicated by distributing labels along the trait scales. The number of labels is usually three or five but it may be more or less than this.

Symonds[1], writing in 1924, reported that the more mature and interested judges, when the trait was a well defined one such as "neatness", could profitably use a scale of seven class intervals, whereas, when the trait was a vague one such as "tact", or when the judges were immature or lacked interest, only five or four class intervals could be clearly distinguished. Symonds approached the question from the standpoint of employing the number of class intervals which would provide maximum reliability by reason of full use of the judge's powers of discrimination without refining the scale beyond the point where those powers would cease to be effective. He accepted as permissible a loss of reliability of 1.099% due to coarseness of the scale. From calculations based on this, he concluded that "In constructing scales for rating traits of personality the optimum number of class intervals is 7."[2] Guilford[3] declared that the optimum number of scale divisions is five, while a more recent discussion by Ghiselli and Brown offered the comment, "The best that can be done is to discover empirically the optimal number of steps for each scale."[4]

to further than I do. It is common practice in building guidance for the rater in his choice of his mark to be checked on each trait scale. Varying degrees of possession of the traits are indicated by distributing lines along the trait scales. The number of lines is usually three or five but it may be more or less than this.

Yoder, writing in 1941, reported that the more serious and interested judges, when the trait was a well defined one such as "neatness", could practically use a scale of seven clear intervals, whereas, when the trait was a vague one such as "race", or some "judge" was interacting or lacked interest, only five or four clear intervals could be clearly distinguished. Yoder approached the question from the standpoint of analysis, the number of cases involved which would provide sufficient reliability by reason of full use of the judgments of discrimination elicited in filling the scale beyond the point where more power would cease to be effective. He proposed as sufficiently a line of reliability of .0007 due to consciousness of the scale. Experimental data seem on this, he concluded that "in constructing scales for rating traits of personality, no problem number of clear intervals in a 7-point scale" [illegible] [illegible] number of [illegible] [illegible] [illegible]

The space on the trait scales between labels may be graduated to permit finer distinctions between ratees. However, the value of these smaller subdivisions of the scales is deprecated by many for the same reasons given above for not using too many class intervals.

The labels used may be merely evaluative terms such as "excellent", "good", and "fair", but the preferred practice is to employ descriptive phrases which define what is excellent, and what is good, and what is only fair. This conforms with the findings of experimenters, to be discussed more fully later, that better results are obtained on rating scales in general which make use of descriptive rather than evaluative terms.

Scoring the Graphic Scale

Scoring of the markings on the graphic scale is usually the responsibility of the central personnel agency. It may be accomplished in any of several ways, the most common of which will be referred to as, the profile method, the direct calculation method, the graphic-computation method, and the machine method.

<u>The profile method.</u>[5] In this plan the markings are converted into a profile by simply connecting by straight lines the points checked on each trait scale or by constructing a similar figure on a specially prepared profile sheet. Such graphic presentations of the individual scores of all the ratees may be compared with standard profiles and the individ-

The sponge on the trail section [illegible] in its way be productive to put in first [illegible] [illegible] [illegible]. However, the value of these [illegible] qualifications of the [illegible] is depreciated [illegible] the [illegible] [illegible] not using too many [illegible].

The [illegible] [illegible] [illegible] as "[illegible]", "[illegible]", and "[illegible]", [illegible] [illegible] [illegible]

uals may be ranked in accordance with the degree with which their profiles approach the standard.

Profiles are useful in determining the suitability of an individual for a position for which the most satisfactory profile has been determined. However, their use would not be practicable when comparing the ratings of large numbers of individuals, or when any of a number of profiles would be satisfactory for the job, as would probably be the situation in the rating of Naval officers. The writer knows of only one attempt to determine the most satisfactory profile or profiles for Naval officers. It is being carried on in connection with the leadership studies being conducted by The Ohio State University under contract with the Department of the Navy. These studies will be discussed in their bearing on "criteria" in the next chapter.

<u>The direct calculation method</u>. Rating systems using this method provide the scale values, scoring weights, and space for computation and tabulation directly on the rating blank. It is a little used plan, however, because most rating procedures involve scoring in a central office where the scale values and scoring keys can be conveniently furnished separately, thus simplifying and reducing the cost of the rating blanks. An additional consideration is the desirability of omitting from the rating blanks any indications of the relative values of the individual traits to avoid the possibility of their influencing the rater.

[illegible] may be ranked in accordance with the degree of [illegible] their profiles approach the standard.

Profiles are used to determine the suitability of an individual for a position for which the most satisfactory profile has been determined. Hence, [illegible] not be practically [illegible] the [illegible] of individuals, or even any of a [illegible] of profiles, [illegible] satisfactory for the job, [illegible] the testing of naval officers. [illegible] attempt to determine the most satisfactory profile or profiles for naval officers. It is [illegible] the Isaacsohn study [illegible] naval officers with varsity junior officers [illegible] the navy. These studies will be discussed in greater detail in the next chapter.

<u>The direct method of rating</u>. This method of rating [illegible] this method provide the male [illegible] space for comparison and to [illegible] interest, on the rating blank. It is a little used [illegible] [illegible] ing procedure involve only [illegible] [illegible] are merely [illegible] [illegible] [illegible]

The graphic-computation method. In this procedure the marks on the individual traits are transformed into numerical values by measuring along the scales in millimeters or other suitable units. These individual trait scores are combined, in accordance with the scoring key, into an overall score. Burtt[6] described this method and a refinement of it in which stencils aid in picking off the scale values.

When scoring ratings in large numbers, manual routines become time and energy consuming to an excessive degree. This detracts seriously from their usefulness to the Navy.

The machine method. When large numbers of ratings are processed, scoring by the International Business Machines Corporation's scoring machine is not only quick and accurate, but also economical. In order that this system may be used, the marks must be placed on special scoring sheets, in the small spaces provided, using special "electrographic" pencils. As the sheets move through the machine at the rate of about 500 per hour, electric brushes pass over them and register an electrical impulse each time they touch the graphite of the pencil marks. These impulses actuate the scoring and tabulating mechanisms of the machine which produce the sought after scores. The special scoring sheet may be separate from the rating blank or, as in the case of the Army's new rating, shown in Appendix E, the two may be printed on the same piece of paper.[7]

The graphic-correction method. In this procedure the marks on the limited well logs are transformed into material values by measuring along the scales in millimeters or other suitable units. These individual tracks across are combined, in accordance with the correct key, into an overall score. Usually [illegible] this method of a [illegible] of it in which samples are [illegible] off the [illegible] values.

When working with a [illegible] large number, [illegible] formations become time and energy consuming to an excessive degree. This [illegible] arises from [illegible] inefficiency in the heavy.

The machine method. When large numbers of [illegible] are processed, [illegible] by the [illegible] machine. Commercially [illegible] machine is not only quick and accurate, but also economical. [illegible] the [illegible] machine is placed on [illegible] [illegible], in the small square [illegible], using [illegible] "[illegible]" [illegible]. As the [illegible] [illegible] [illegible] [illegible] 500 per hour, [illegible] [illegible] [illegible] an [illegible] [illegible] [illegible] [illegible] the [illegible] [illegible]. [illegible] [illegible] [illegible] [illegible]

Types of Scores

The overall raw scores obtained by any of the foregoing methods can be presented in any of several forms. They may simply be recorded as numerical raw scores or they may be translated into letter scores such as A, B, C, D, or E, etc. Other possibilities are percentile scores, standard or z scores or simply rank order positions.

The numerical score form is not a highly regarded end result. Authorities generally feel that the precision presently attainable in rating is not sufficient to give meaning to differences in numerical scores unless those differences are relatively large. Tiffin[8] warns against recording total ratings numerically in values covering a wide range and adds that it is better to use ratings of A,B,C,D, and E than to try to explain why one man is rated 248 and another 249 on a scale covering a range of 250.

Ranking is a simple procedure and is easily understood by everyone, but it suffers from one important shortcoming. Burtt calls attention to this weakness with the words, "There is nothing to indicate whether the steps between successive pairs of ranks are equal or otherwise,....".[9]

The percentile score and the z score have much to recommend them. The former is readily understood and, in a sense, any two percentile scores are comparable. However, one must remember that a difference of five percentile units near the middle of the distribution of scores reflects less

Types of scores

The overall rhythm achieved by any of the forming methods can be presented in any of several forms. They may simply be recorded as numerical rhythm scores or they may be transcribed into letter notation such as A, B, C, D, or E, etc. Other possibilities are pictorial scores, shapes or a scores or simply rhythm-tone patterns.

The numerical score form is not a highly regarded and usual. Auditoriums (usually that which is available) are easily obtainable in metric, is not sufficient to give meaning to differences in numerical scores which are differences are relatively large. Giving scores without according total points numerically in values covering a wide range and that it is better to use ratios of 1/2/3, and 4 when to try to explain why one has 1/2 and another 3/3 or a scale covering a range of 200.

Meaning in a single procedure can be easily misunderstood by everyone, but it suffers from an important shortcoming. Much calls attention to this weakness with its warning: "There is nothing to indicate whether the line between successive pairs of scores are equal or otherwise."

The pictorial form is a kind of graph. It is [illegible] [illegible] [illegible]

difference in ratings than the same difference in percentile units at either end of the distribution. Though less easily understood by the person who is not familiar with the science of statistics, z scores are free from this difference in value between the middle of the distribution and the ends. Further, all z scores are fully comparable with each other and they have added utility in that they can be averaged.

Weighting of Traits

It is obvious that some qualities are more important to success in any particular job than are others. Because of this it is usually desirable to adopt a system of scoring in which the marks on the more important traits are made to "count more" in the overall score than those on the less important items. This weighting is accomplished by multiplying the marks on selected traits by numbers (weights) before computing the overall score. Usually the weights are integers greater than unity.

The magnitudes of these multipliers are determined by expert judgement. Persons experienced in the requirements of the job for which a rating system is being designed, are requested separately to assign what they consider to be the proper weights within a prescribed range of values. The medians of such judgements, normally "rounded off" to the nearest integral value, become the official scoring weights. Marble described the use of this procedure in an actual de-

difference in feeling than the same difference in pronunciation unless it alter one of the syllables. This is less easily understood by the person who is not familiar with the solecism of accentuation, a course arising from this difference in values between the ability of the distinguish and the ear. Furthermore, all a novice can fully appreciate with ease only r and they have added utility in that they can be averaged.

Relevance of Drama

It is obvious that four qualities are more important to readers in any particular. Yet some are minor. Because of this it is usually necessary to make a choice of scoring in which the marks on the more important traits are made to "count more" in the overall score than those on the less important items. This adjustment is accomplished by multiplying the marks on minor traits by number (usually) before combining the overall score. Usually the weights are integers greater than unity.

The importance of these multipliers are determined by expert judgement. Persons experienced in the requirements of the job for which rating scales are being constructed, are frequently knowledgeable in spite of bias. Consider to be the proper weight, if it is a particular trait or feature. The number of ways in which a particular trait or feature may be used in the job, number of times it is used, usually, number of times it is used in important tasks, number of times it is used in critical situations, and the degree to which it is used in critical situations.

sign situation.[10] Another convenient method involves having the experts list the job's principal traits in the order of their importance. From these nominations weights may be derived, based on the frequency with which each trait is mentioned. A third method was used by Knauft[11] whose group of experts used the "Equal Appearing Intervals" technique of L. L. Thurstone in choosing their weights. It is readily apparent that in each of the three foregoing procedures, assignment of weights is based on the "face validity" of the traits being considered.

Burtt[12] offered a fourth method of determining scoring weights based on the reliability with which the traits may be judged,[13] that is, the extent of concurrence between two judges in estimating the same trait or the degree of agreement between successive estimates of the same trait by the same judge. He argued that a highly reliable trait may be given a heavy weight on the assumption that, since it is difficult to determine validity, it is better to deal with the more reliable traits. Moreover, if the judges can agree more closely among themselves on some traits than on others, the former should be given more weight not because they relate more closely to proficiency on the job[14] but because the ratings themselves come nearer to being a true rating of the more reliable traits.

The writer questions the desirability of assigning weights on this basis. Undoubtedly it will lead to improve

sign alternation.[10] Another convenient method involves making the experts list the judge's principal traits in a order of their importance. From these rankings weights may be derived, based on the frequency with which each trait is mentioned. A third method was used by Hunter[11] whose group of experts used the "Equal Appearing Intervals" technique of L. L. Thurstone in ascribing trait weights. It is probably apparent that in each of the three foregoing procedures, assignment of weights is based on the "face validity" of the traits being considered.

Ferguson[12] offered a fourth method of determining weights based on the reliability with which the traits may be judged,[13] that is, the extent of correspondence between two judges in ascribing the same trait to the ratee of agreement between successive judgments of the same trait by the same judge. He argued that a highly reliable trait may be given a heavy weight in the composite score, since it is difficult to discriminate well in it, it is better to lump all the more reliable traits. However, if one judge gives a very [illegible] rating on one trait while the other gives a very [illegible] rating on the same trait, since it is the [illegible] which [illegible] [illegible] [illegible]

the reliability of the rating device, but, although validity is dependent upon reliability, heavily weighting an unimportant trait seems likely to undermine rather than bolster the total validity of the rating. Consider a hypothetical extreme situation in which a trait which can be estimated with great reliability (for instance, with a coefficient of reliability of +.95) but which has absolutely no relation to proficiency on the job, has been included on the rating form. What possible value has it, no matter what weight is assigned?

Need for Control of Weighting

Unless special precautions are taken, the different traits may become weighted in a manner not intended. "If you are not considering the variability of the ratings on each trait," wrote Tiffin and Musser, "the trait of least important actually may be receiving the heaviest weight... ...When combining scores-- regardless of the nature of them-- they weight themselves automatically in proportion to their respective variabilities-- standard deviations."[13] Tiffin[14] subsequently discussed this phenomenon more fully and gave examples showing how the standard deviation affects the weighting, even to the extent that a completely inaccurate overall rating may result.

The authors[15] recommend control of weighting by transforming trait raw scores into z scores (standard scores) which, as they explain, may be multiplied by the weights without fear of inaccuracy resulting. These weighted z scores may then be

the reliability of the total test s. but, although validity is dependent upon reliability, locally weighted, or important test items seems likely to contribute rather than detract the total validity of the total. Another approaches extreme allocation in which a trait factor may be measured with great reliability (for instance, with a coefficient of reliability of .90) but which is least reliable relative to proficiency on the job, may seem reasonable on the basis of what possible value it is, no matter how small it is analyzed.

Need for Control of Criterion

Unless a careful research on the criterion, the different traits may become well defined in a manner not intended. "If you are not considering the variability of the weights on each trait," wrote Tiffin and Cleaver, "the trait of least important actually may be receiving the heaviest weight... ...The conclusion seems--regardless of the method--they weight therefore relatively in proportion to their respective variabilities--another variation."[13] Tiffin[14] subsequently illustrated it in experiment for Ill. per were exactly alike the standard deviation of effect, the well-fitted, [illegible] complete. Otherwise overall validity [illegible] result.

The [illegible] measure [illegible] [illegible] [illegible]

totaled for each individual to obtain his overall standard score.

A simpler device for correcting scores and thereby controlling weighting was suggested by Guilford[16] in the form of a scoring stencil with divisions spaced unevenly. The spacing would be based on the difference between a normal distribution and the average of the actual distributions of the scores.

Advantages of Graphic Scale

The chief advantages of the graphic rating scale may be briefly stated as follows. First, it lists all of the traits which are important to the job and thereby insures consideration of each. Second, it is easily understood by raters and ratees alike. Third, if it is not too lengthy, it may be filled out and used without an excessive expenditure of time and energy. Fourth, it is relatively easy to score by hand and is easily adaptable to machine scoring. And fifth, it provides a specific picture of the ratee's strong and weak points which is useful to the ratee and to management, as has already been discussed under "OBJECTIVES OF RATING" on pages 2 and 3.

Disadvantages of Graphic Scale

The graphic scale, though extensively used because of its important advantages listed above, is subject to nearly all the errors of a psychological origin that are likely to occur.

<u>Failure to obtain spread of scores</u>. The most readily observed shortcoming is the failure of the raters to produce

totaled for each individual to obtain his overall attained score.

A simpler device for correcting response and thereby controlling weighting was employed by Wilfrong[10] in the form of a scoring stencil with circles opened overlay. The stencil would be based on the difference between a normal distribution and the average of the actual distributions of the scores.

Advantages of Graphic Scale

The chief advantages of the graphic scale may be briefly stated as follows. First, it lists all of the traits which are important to the job and it gives increased consideration of each. Second, it is easily understood by raters and raters alike. Third, it is quickly tabulated. It may be filled out and used without an excessive expenditure of time and energy. Fourth, it is relatively easy to score by hand and it easily adaptable to machine scoring. The fifth, it provides a specific picture of the raters attitude and weak points which is useful in the rater on the same basis, as has already been discussed under "Advantages of Ranking" on pages 2 and 3.

Disadvantages of Graphic Scale

The use of the scale, on the other hand, has certain disadvantages. It is important a variance [illegible] all the errors of a personal lost [illegible] to [illegible] .

Failure to [illegible] of [illegible]. [illegible] may [illegible] [illegible] in the failure of [illegible] to [illegible]

a spread of rating scores. That is to say, they fail to obtain anything approximating a normal distribution of scores. Thorndike wrote, "...the variations in any single trait are usually continuous. ...the variations usually cluster around one and only one type."[17] He was referring to, (1) the fact that most people will show a trait or a quality to an average degree or very nearly so, while a few other people will be above or below average, and only a very few will be greatly above or below average and, (2) there are no steps or classes but rather, a continuum in the distribution. The mathematical expression of this distribution is the Gaussian "bell shaped" curve. A through treatment of the characteristics of this curve may be found in Adkins[18] or any standard statistical text. Psychological literature is replete with further discussion and accounts of experimental demonstrations in support of the normal distribution theory. Similarly, most discourses on service rating devices and procedures advance the view that properly made service rating scores will rather closely approximate a normal distribution.

An interesting exception to the unanimity described above may be found in the discussion of service ratings by Mosher and Kingsley.[19] These writers feel that the so called "J curve" type of distribution is to be expected for merit rating scores and, moreover, is quite proper. Their arguments are based on claims that the behaviors recorded are institutional behaviors and the population dealt with is a selected

one.

In saying that there is a failure to produce a spread of scores, reference is made to the tendency of raters to mark all ratees high. This tendency results in a skewed distribution which, if we follow the consensus of psychological opinion, disagrees with reality and makes necessary special procedures to prevent uncontrolled weighting, as was mentioned earlier.

The problem of reliability. The very form of the graphic rating blank tends to increase the reliability of the rating. By requiring the separate consideration of each trait, the tendency to rate on the basis of a recent event, an attitude, a preconceived notion, or halo, is minimized to some extent. It is generally agreed that the resistance of the graphic form to halo effect may be increased by arranging the individual trait scales so that the favorable labels alternate irregularly from the right to the left ends of the scales. Jucius[20] recommended also elimination of all graduations from the trait scales (leaving only the labels) and spacing these labels unevenly. These measures will force the rater to read the labels on each individual trait scale and, it is reasoned, consider each trait more carefully before marking. Burtt[21] and Guilford[22] suggested that the blank be designed to cause the rating of all ratees on one trait before the next trait is considered. For any considerable number of ratees this would require a separate sheet for each ratee for each trait.

one.

In saying that there is a "risk" to produce a spread of access, reference is made to the tendency of radio to mark all crises sign. This can only results in a biased television which, if we follow the concept of psychological opinion, disposes with reality and makes necessary special procedures to prevent uncontrolled reactions, or the manipulated reality.

The problem of reliability. The very fact of the dramatic radio, alone take, to increase the reliability of the rating. By replacing the imperfect description of each item, the tendency to rate on the basis of a recent event, dramatic task, a preconceived notion, or also, "a misleading halo" extent. It is generally agreed that the weakness of the graphic have to this effect and - increased by using the individual ratings rather than a rankings. Usually, after data [illegible] to [illegible] [illegible] [illegible]

The obvious objection to such a plan is its cumbersomeness.

Another practice which will bring the rater to more thoughtful consideration of the rating is to require that each trait mark be substantiated by a brief account of the ratee's behavior on which the mark is based. A variation of this plan used by some, provides for a notation as to whether the individual trait mark was based on actual behavior or on subjective judgement.

The rating system used in the U. S. Air Force for officers (and described more fully in a succeeding chapter) strives for accuracy by using a standard booklet for each ratee in which day by day observations of significant behaviors are recorded by check-marking in appropriate spaces. At the end of the period when the formal rating is to be made, the consensus of the running record is transferred to a graphic scale sheet which becomes the official rating.

Errors caused by failure of the rater to understand the meaning of the trait names may be reduced by substituting for them phrases descriptive of the behavior which will indicate the presence of those traits. This will also tend to reduce the difficulty due to lack of knowledge by the rater as to what overt behavior he may take as an indication of a particular trait. The superiority of descriptive phrases over trait names was demonstrated by Spockford and Bissell[23] experimentally. A clue as to the reason for this superiority may be found in the writings of Kerble[24] who concluded that,

The obvious objection to such a plan is the concentration.

Another practice which will bring one more to more thoughtful consideration of the matter is to require that each trait must be substantiated by a brief account of the rater's behavior on which any such is based. A modification of this plan used by Kohs, required for a rating on a worker the individual traits were used on actual behavior or on subjective judgment.

The rating phrase used in the U. S. Employment Service (and described more fully in a subsequent chapter) strives for accuracy by making a minimum possible for each rater in which one by the elimination of significant behaviors are measured by observation. In approximate success. At the end of the period when the formal rating, in its in terms, the behavior of the worker is transferred to a graphic media chart which records the official rating.

Errors resulting by failure of the rater to differentiate the ability of the trait rather than by reduced by one difficulty from other different traits of the individual will indicate a certain degree of error. This will lead to the rater the difficulty in the line of the [illegible] as to the other [illegible] [illegible] [illegible]

"While it may be and frequently is a matter of considerable debate whether an employee is 'resourceful' or not, it may be relatively more simple to agree on whether or not the employee engages in certain observable activities which are by agreement a component part of resourcefulness."

The problem of validity. Validity is sought by, (1) using traits which relate closely to the job and which are important to success in it and, (2) by making the rating device and procedure as reliable as possible. Criteria against which to check validity are sadly inadequate in most cases, as will be discussed in the next chapter. Consequently psychologists are denied the luxury of selecting items empirically and must rely on the judgement of experts for this function. Their hypothesis is that if items which are judged to be valid (and which therefore must appear to relate closely and importantly to the job) can be reliably rated, the overall rating will be valid. Cronbach's declaration that, "Observations by impartial observers are generally accepted as valid if they can be made reliable,"[25] illustrates this view. A more precise statement of the relation between reliability and validity, and one which further emphasized the importance of the former to the latter, was given by Adkins who wrote, "Thus it is clear that, except for chance factors, the validity coefficient of a test cannot exceed the square root of its reliability coefficient."[26]

FORCED DISTRIBUTION RATING

Attempts to secure an approximation of the normal distribution have led to the use of the "Forced Distribution" in connection with the graphic scale as shown in Appendix A. For this technique a number of categories from best to poorest is established and the percentage of ratees who may be placed in any category is prescribed. When rating, each ratee is considered in terms of a single trait and then placed in the category deemed proper. When all have been rated and categorized in one trait, the process is repeated for the remaining traits in turn. An overall score may be computed from the ratings thus performed. Tiffin suggested the following distribution for a rating of this type:[27]

Category	Percentage of ratees in the category
Highest	10
Next highest	20
Middle	40
Next lowest	20
Lowest	10
Total	100

As shown in Figure 1, page 37, this distribution approximates the bell shaped curve of the normal distribution.

If conscientiously executed, the forced distribution system will tend to give a more valid spread of scores, provided a large number of ratees are rated by the same rater. However, where the number is not large, it is probable that the merit of the employees under any one supervisor will not

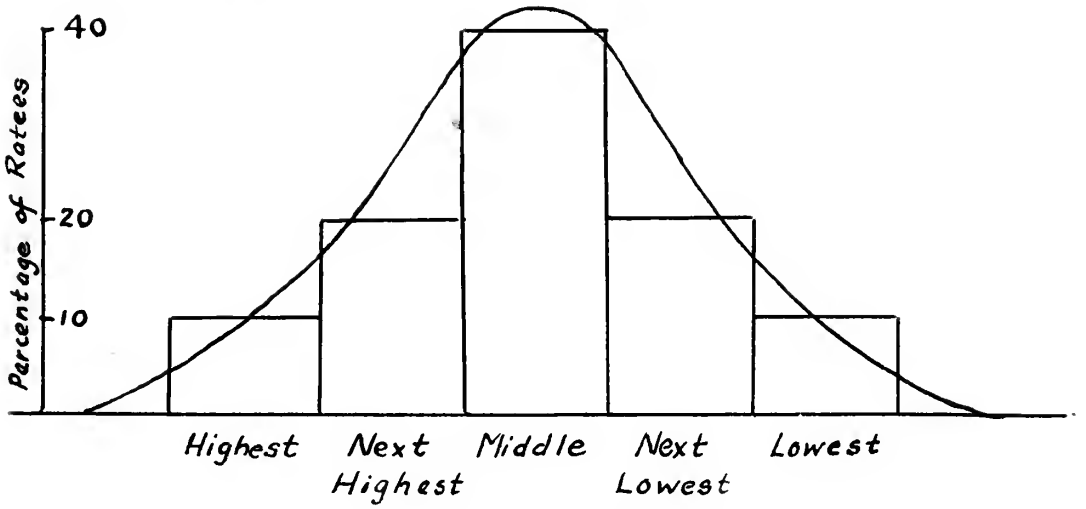
RATE - LEVEL TIME ANALYSIS

Attempts to secure an approximation of the normal distribution curve for the "Forward Interpolation" in connection with the Jewish scale as a new interpretation, this technique a number of preliminary tries must be produced in data linked and the problems of testing [illegible] be placed in any category of [illegible]. The results were noted to correlated in the order of a single scale [illegible] from Passed in the category based process. Then all have been [illegible] and [illegible] quoted in the order, the process is [illegible] the remaining tunics in [illegible]. In overall [illegible] [illegible] the results [illegible]. Rates [illegible] the "allowing distribution for a [illegible] of this [illegible]:[27]

Category	Percent of Cases (in the [illegible])
Highest	10
Best [illegible]	20
Middle	40
Less Jewish	20
Lowest	10
Total	100

As shown in Table 1, [illegible] 30, [illegible] interpolation [illegible] the [illegible] of the normal [illegible] curve.

If [illegible] Rate/Level [illegible] [illegible] [illegible]



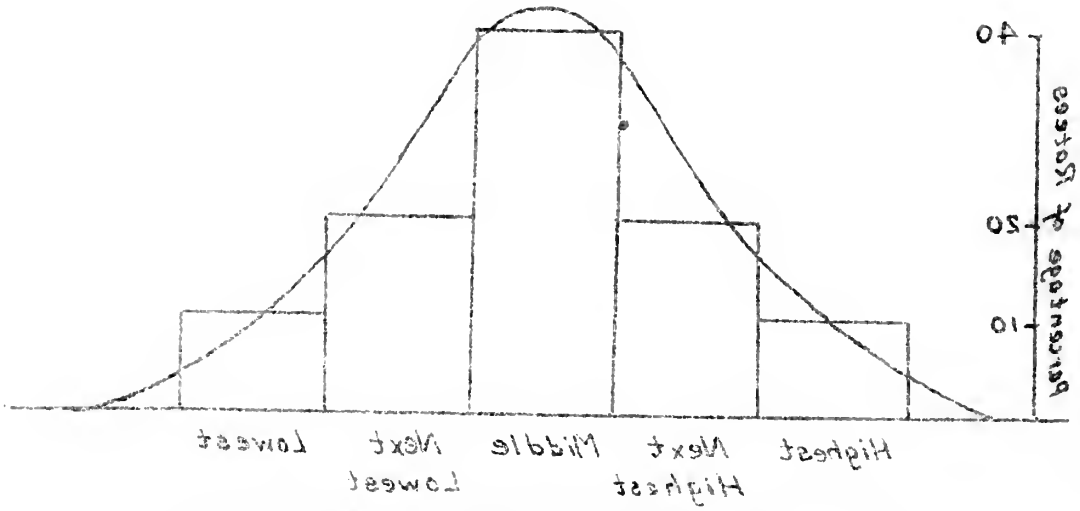
Categories of Prescribed Distribution

Figure 1

A DIAGRAM SHOWING HOW PRESCRIBED DISTRIBUTION APPROXIMATES NORMAL DISTRIBUTION

cover the entire range from best to poorest when compared with the merit of all the other employees. If such were the case, a skewed distribution for the ratings of one or more of the supervisors would be proper and any plan to force normal distributions would work an injustice.

The assignment of Naval personnel is not entirely a randomized operation, particulary when detailing officers, and it is probable, therefore, that there will be differences in merit of the officers in one command from those in another. Further, the program of rotation of duty frequently changes the membership of groups of officers. Because of this, any plan of forced distribution is unsuitable for a Navy rating system.


Percentage of respondents
10
20
40
Highest
Next
Middle
Next
Lowest
Lowest
Highest

RANK ORDER METHOD

It is known that raters find it relatively easier to arrange their subordinates in order from best to poorest than to rate them using the adjectives or descriptive phrases of the usual graphic scale. This procedure of ranking is used as a formal rating procedure in some cases and is known as the "Rank Order Method".

Ranking may be concerned only with a single overall quality such as "job performance" or "promotability", or it may be performed on each of several traits in succession. It is not afflicted with the error of the skewed distribution since it presents the ratees only in the order of their excellence relative to each other without regard as to how many are actually "excellent", "average", or "poor".

This system suffers from the same disadvantages and unsuitableness for Navy use discussed for the Forced Distribution System. In addition, like all systems of ranking, this plan is based on the erroneous assumption that the difference between any adjacent pair of ranks is equal, or may be treated as equal, to the difference between any other adjacent pair (see passage quoted from Burtt on page 27).

PAIRED COMPARISONS

It is relatively easy to choose which of several men is the best and which is the poorest but, as one approaches the mean of the distribution, discrimination becomes more difficult. An aid in this situation is the method of "Paired

[illegible]

It is known that [illegible] is [illegible] to
arrange [illegible] [illegible]
to note [illegible] [illegible]
the [illegible] [illegible] [illegible]
as a [illegible] [illegible] [illegible]
the "[illegible]".

[illegible] [illegible] [illegible]
quality [illegible] "[illegible]" [illegible] "[illegible]", [illegible]
may [illegible] [illegible]
is not [illegible] [illegible] [illegible]
since [illegible] [illegible] [illegible]
[illegible] [illegible] [illegible]
are usually "[illegible]", "[illegible]", [illegible] "[illegible]".

[illegible] [illegible]
[illegible] [illegible] [illegible]
tion [illegible]. [illegible] [illegible]
place [illegible] [illegible]
[illegible] [illegible]
as [illegible] [illegible] [illegible]
(see [illegible] [illegible] [illegible] [illegible]).

[illegible]

[illegible]

Comparison" in which each individual is compared in turn with every other, in the group being rated, and a judgement is made at each comparison as to which is the better. From these judgements a reliable rank order can be constructed. As in the Rank Order Method, rankings may be obtained on a single overall quality or on a series of specific traits. An application of this method to an actual rating situation was described by Lawshe, Kephart, and McCormick.[28] Ratings were obtained with a reported average reliability coefficient of .83.

The method of Paired Comparisons has an additional advantage in that it offers a means of determining the relative values of the differences between adjacent pairs of ranks. Thurstone[29] and later Guilford[30] described in detail the process involved. It is a workable and satisfactory procedure for scaling items for use on an attitude scale but is likely to require too much time for practical use in scoring ratings. Furthermore, the product of this method of rating is an order of ranks, which, as we have discussed, shares with the Forced Distribution, a real risk of injustice to certain ratees.

There is a serious disadvantage attached to the Paired Comparison Method. Guilford[31] commented that it takes too much time and is "wearying" to the judges, and reported that a German investigator, L. Witmer, in 1894 criticized the method, after trial, also on the grounds that "it took too much time". A subsequent group of investigators have express-

comparison" in which each individual is compared in turn with every other, in the group being rated, and a judgment is made as each comparison as to which is the better. From these judgments a complete rank order can be constructed. As in the Rank Order Method, rankings can be obtained on a single overall quality or on a series of specific traits. An application of this method to an actual rating situation was described by Lawshe, Kephart, and McCormick.[28] Rankings were obtained with a reported average reliability coefficient of .88.

The method of pairs comparisons has an additional advantage in that it offers a means of determining the relative values of the differences between adjacent pairs of ranks. Thurstone[29] and Lazer Guilford[30] described in detail the procedure involved. It is a variant, and satisfactory procedure for scaling items for use on an attitude scale, but is likely to require too much time for practical use in scaling ratings. Furthermore, the product of this method of ranking is an order of merit, which, as we have discovered, clashes with the Forced Distribution, a real risk of injustice to certain ratees.

There is a scales itself-rating approach to the Forced Comparison Method. Guilford[31] recommends that it seems too much like one is "keeping" to the judges, and reported that a German investigator, Dr. Wagner, in fact controlled the method, after trial, also on the grounds that "it took too much time". A modification, using a three-point scale, e-

ed the contrary opinion, however.[32]

Mathematicians have shown that when any certain number of objects are to be paired each with every other in the group, the number of separate pairings is given by the formula:[33]

$$\frac{n(n - 1)}{2}$$

where n is the total number of objects to be compared. Thus if only 10 ratees are to be rated by this method, the number of separate comparisons will be:

$$\frac{10(10 - 1)}{2} = 45$$

If comparisons are to be made on more than one quality, the total number of comparisons will be increased accordingly.

Despite the technical excellence of the Paired Comparison Method, its cumbersomeness prevents wide use of it and, coupled with its inherent risk of injustice to certain ratees, renders it unsuitable for Navy use.

MAN TO MAN SCALE[34]

The Man to Man Scale was developed by the Bureau of Salesmanship Research[35] and put to extensive use by the U. S. Army in 1917. It was discarded in 1920 for the graphic scale and is mentioned here only because it served as the prototype from which the graphic scale was developed.

The Man to Man device required that for each trait or quality, the rater set up a master scale listing the names of actual persons known to him arranged in the order of their

ed the military opinion, however.[32]

Mathematicians have shown that when only certain number of objects are to be paired each with every other in the group, the number of separate relations is given by the formula:[33]

$$\frac{n(n-1)}{2}$$

where n is the total number of objects to be compared. Thus if only 10 teams are to be paired by this method, the number of separate comparisons will be:

$$\frac{10(10-1)}{2} = 45$$

If comparisons are to be made on more than one quality, the total number of comparisons will be increased accordingly.

Despite the technical deficiencies of the Guilford Zimmerman method, it commandeered immediate wide use of it and, coupled with its inherent view of inclusion to certain values, rather is unable to ever "die" out.

MAX TO THE CRAZE[34]

The Max to Man test was developed by the bureau of educational research[35] and was so extensive use by the U. S. Army in 1917. It was discarded in 1920 for the graphic scale and is mentioned here only because it marked the change type from which the graphic scale was developed.

The Max to Man device consisted then to each trait or quality, the rater had to a number each different on the basis of actual observed behavior to the rater's or ratee's

excellence in the trait. Rating consisted of comparing the ratee with these master scales and assigning him ranks or scores equivalent to those of the persons on the several master scales whom he resembled the most closely in traits.

The task of constructing these master scales proved too laborious and this led to the adoption of the graphic scale which substitutes for the comparison of the ratee with a master scale of actual people, a comparison with a scale of standardized traits or descriptions. This difficulty due to the master scales is a real disadvantage in the practical situation. Because of this and because there are better ways of rating one's subordinates, the Man to Man Method is not suitable for Navy use.

CHECK LIST

In 1927 Probst[36] began research to develop a new type of rating. The result has been described as a "check list" because of its form. As shown in Appendix A, it presents a series of descriptive statements with provision for the rater to check-mark those which are applicable to the ratee. Scoring weights have been determined by experiment for each item and the sum of the weights of the items checked, when placed in the scoring formula (also developed by "trial and error")[37] yield a numerical score which is converted by use of a table into one of ten letter grades.

A number of trials of the device using populations ranging from N = 20 up to N = 3039 show a consistent ability

[illegible] in the trail. [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] equations [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]

to produce distributions which approximate normal distributions rather well.[38] Reliability coefficients of +.78 when N was 475 and +.918 for a smaller group were reported[39] while validity coefficients, with supervisor judgements as criteria, range from +.36 to +.77 for laborers[40] and from +.43 to +.78 for seventy-eight public health nurses.[41] An extraordinarily high validity coefficient of +.922 was reported for thirty municipal probation officers with conference ratings as the criterion.[42]

A report of a more recent use of the Probst type of scale has been made by Knauft. He constructed two similar forms and obtained a reliability of +.87 for one of them[43] and a validity, with supervisors' judgements as a criterion, of +.59 for the other.[44]

In discussing the Probst scale, White wrote, "The evidence indicates that satisfactory results can be secured with it."[45] Mosher and Kingsley also regard the system favorably though they warned of "serious technical weaknesses" in it and discussed them at some length.[46] They criticized particularly the method of choice of scoring weights, the values of scoring weights assigned, and the frank (and apparently successful) attempt of Probst to design a device which will give a nearly normal distribution of scores.

Recently, Probst has explained his choice of scoring weights and method more completely and has reiterated that he did not set out to force his device to render a normal

to produce distributions with an approximate normal distribution rather well.[39] Reliability coefficients of .70 when N was 473 and +.910 for a smaller group were reported.[40] While validity coefficients, with supervisor judgments as criteria, range from +.50 to +.77 for Lawshe[41] and from +.42 to +.78 for several sets of public health nurses.[42] In contrast, relatively high validity coefficients of +.900 was reported for study medical probation officers with conference ratings as the criterion.[43]

A report of a more recent use of the Pretest type of scale has been made by Tuckett. He constructs two similar forms and obtained a reliability of +.87 for one of them[44] and a validity, with supervisors' judgments as a criterion, of +.50 for the other.[44]

In discussing the Pretest scale, Lide says, "The evidence indicates it is a satisfactory technique for measuring skill."[45] Tucker and Lyle also report the method favorably and say, they varied of "making objective evaluation" is and "learned from as time lapses."[46] They criticized particularly the method of choice of scale, [illegible] the values of scaling variable experiment, and the [illegible] apparently [illegible] [illegible] [illegible]

distribution. Instead he reported, "The sole aim was to evolve a scheme that would place an employee in the service group that would be generally recognized as correct for him."47

COMBINATIONS

The never ending search for improvements in rating devices has led to combinations in the one form of two or more methods of rating in the hope that a summation of advantages would result. The addition of the forced distribution provision to the graphic scale has been discussed already. It is usually achieved by dividing the scale into vertical columns each headed by the proper categorical designation as shown in Appendix A. The instructions for the raters using such a form may make it mandatory that the distribution be followed or they may only urge it.

A man to man rating may be included on the combined form. If such is the case, the rater will find instructions to compare the ratee to a list of a prescribed number of individuals known to the rater, and whom he must rank in order from best to poorest. As a result of this comparison, the ratee is to be assigned a rank equal to that of the individual on the list whom he most closely resembles in traits.

Further attempts to obtain validity include requiring the rater to indicate whether, under some specified difficult situation, he would prefer to have the ratee with him, be merely satisfied to have him, or actually prefer not to have him. The subject is also approached from the direction of

distribution. Instead he reported, "The sole aim was to devise a scheme that would place an employee in the service group that would be generally recognized as correct for him."[47]

CONCLUSION

The never ending search for improvements in rating devices has led to combinations in the one form of two or more methods of rating. In the case that a succession of advantages would result. The addition of the forced distribution provision to the graphic scale has been discussed above. It is usually achieved by dividing the scale into vertical columns each labeled by the proper categorical designation as shown in Appendix A. The instructions for raters using such a form may read as follows: "Place the distribution be followed or they may call upon it.

A rater to use this rating may be informed on the combined form. If such is the case, the rater will find instructions to compare the rates to a list of a prescribed number of individuals known to the rater, as when to rank them in order from best to poorest. As a result of this comparison, the rater is to be assigned a name equal to that of a individual on the list which he most closely resembles in traits.

Further attempts to secure reliable judgments of the rater to individuals ratees, under some specified distribution, as well as provide an easy rating scale, may be found artificial and some may be dictatorily perfect but to some it. The judgment of also comparing [illegible] on

requiring the rater to express his feelings regarding promotion for the ratee under each of several hypothetical conditions such as, (1) a large percentage of his peers are to be promoted, (2) a moderate percentage are to be promoted, and (3) only a small percentage are to be so rewarded.

Lastly, a brief essay or overall evaluational statement may be required or invited. In this the rater has considerable freedom to set down anything of importance which may be relevant and which has not been covered sufficiently in the other portions of the rating. This is the most subjective portion of an already too subjective measurement but most authorities feel that it should not be omitted. In addition to eliciting information which might otherwise not be brought out, it has the beneficial effect of giving the rater the feeling that he has not been regimented excessively and has been allowed to express his own opinion in his own way. This is likely to be a more important consideration as rating techniques of a more truly objective character are devised.

POOLING AND REVIEW OF RATINGS

Plural ratings. The well known principle that the resultant of assessments by several judges can be more accurate than an appraisal by a single judge is used in many rating programs. "The reliability and validity of ratings," wrote Guilford, "increase with the number of judges."[48] An example of the effect of increasing the number of judges was reported by Bradshaw as follows:[49]

requiring the rater to express his feelings regarding promotion for the ratee under each of several hypothetical conditions such as, (1) a large percentage of his peers are to be promoted, (2) a moderate percentage are to be promoted, and (3) only a small percentage are to be so rewarded.

Lastly, a brief essay or overall evaluational statement may be required or invited. In this the rater has considerable freedom to set down anything of importance which may be relevant and which has not been covered sufficiently in the other portions of the rating. This is the most subjective portion of an already too subjective measurement but most authorities feel that it should not be omitted. In addition to eliciting information which might otherwise not be brought out, it has the beneficial effect of letting the rater the feeling that he has not been regimented excessively, and has been allowed to express his own opinion in his own way. This is likely to be a more important consideration on rating techniques of a more highly subjective character are desired.

PROBLEMS AND DEFECTS OF RATINGS

Plural ratings. The well known principle that the tendency of assessments by several judges can be more accurate than an appraisal by a single judge is used in many rating programs. "The reliability and validity of ratings," wrote Guilford, "increase with the number of judges."[48] An example of the effect of increasing the number of judges was reported by Bahanan as follows:[49]

Trait	Coefficient of Reliability---	+.75	+.80	+.85	+.90	+.95
1	No. of raters -	17	22	32	50	108
2	No. of raters -	3	5	7	10	21
3	No. of raters -	5	7	10	16	33
4	No. of raters -	7	9	13	21	44
5	No. of raters -	5	6	9	14	29

Such plural judgements may be made and combined in conference into a single appraisal or they may be made separately, without consultation, and then averaged. Proponents for either procedure can be found among the authorities in the field of rating.

Favoring the conference plan are both MacCullough[50] and Yoder.[51] The latter exhibited the extent of his enthusiasm by urging rating by committee even if necessary, because of lack of additional supervisors with the requisite knowledge, to include as members of the rating committee fellow employees of the ratees.

On the other hand, both Tiffin[52] and Burtt prefer to secure the advantage of pooled ratings by averaging ratings made individually, and without discussion. Burtt wrote:[53]

> It has been shown in various connections that greater validity is obtained by averaging independent estimates than by having judges sit together as a committee and make a joint estimate.

Elsewhere[54] Burtt reported that in a study of pooled judgements, when the judgements of twelve judges were each correlated with the criterion the average of the twelve correlations was +.37. However, when the same twelve judgements

Trial	Coefficient of Malleability					
1	No. of raters =	17	22	31	30	100
2	No. of raters =	5	7	7	10	29
3	No. of raters =	5	7	10	10	32
4	No. of raters =	7	5	15	19	46
5	No. of raters =	5	5	5	14	29

such plural judgments may [illegible] and combined in conference into a single appraisal or they may be made separately, without consultation, and then averaged. Proponents [illegible] proceedure can be found among the authorities in the field of rating.

Favoring the conference idea are both Thorndike[80] and Yoder.[81] The latter emphasizes the merits of the extreme by using rating by committee from 12 overseers, because of lack of additional supervisors, with the negative knowledge, so it looks as though in the [illegible] committee follow employees of the release.

On the other hand, [illegible][82] and [illegible] prefer to secure the answers of pooled raters by [illegible] rating made individually, and without consultation. [illegible] says:[83]

> It has been shown [illegible] opinions contributing that greater reliability is obtained by averaging the judgments of raters than by having judges sit together, [illegible] and make a joint estimate.

Flanagan[84] [illegible] that is a mixed of pooled judgments, [illegible] conference of [illegible] judges who [illegible] with the estimates [illegible] [illegible] it as [illegible]. However, [illegible] joint judgments

were combined into a single figure, this composite judgement correlated +.46 with the criterion.[55]

Probst's scale and accompanying instructions indicate that their designer agrees with Tiffin and Burtt. Spaces are provided opposite each item on the scale, for check-marking by three raters. The directions promise that better results will be achieved if the second and third raters will cover the columns containing the check-marks of preceding raters and perform their own ratings uninfluenced by those marks.

Review of ratings. Review at higher echelons of ratings made at lower levels is another possibility for improving judgements. As mentioned above, the Probst scale provides for rating by three persons, with the instructions recommending that the raters make their judgements in inverse order of their seniority. Although, as explained above, it is intended that each perform his task innocent of any knowledge of what his predecessors have done, obviously the knowledge that the ratings performed by juniors will be compared with ratings made by their superiors will have the effect of "review at higher echelons".

The U. S. Army, the U. S. Air Force, and the U. S. Civil Service each follow the practice of review at higher levels. Spaces are provided on their respective rating forms for the action and signature of the reviewing official. The Civil Service in addition requires that ratings be reviewed and approved by a committee.

were combined into a single figure, this composite is found correlated +.46 with the criterion.[58]

Probst's scale and accompanying instructions indicate that their designer agrees with Tiffin and Burtt. Spaces are provided opposite each item on the scale, for check-marking by three raters. The directions indicate that better results will be achieved if the second and third raters will cover the columns containing the check-marks of preceding raters and perform their own rating unaffected by those marks.

Review of ratings. Values at higher echelons of rating made at lower levels in evaluating capability for improving judgements. As mentioned above, the Probst scale provides for rating by three persons, with the instructions recommending that the raters base their judgements in inverse order of their seniority. Although, as explained above, it is intended that each performs his task innocent of any knowledge of what his predecessors have done, obviously the knowledge that the rating performed by juniors will be compared with ratings made by their superiors will have the effect of "review or higher echelons".

The U. S. Army, the U. S. Air Force, and the U. S. Civil Service each follow some variation of rating at higher levels. Spaces are provided on check-marking rating forms for the action and signature of the reviewing officer. The Civil Service in addition requires the rating officer and reviewer to be specific regarding the strengths and weaknesses of a candidate.

When preparing to implement a new rating device,[56] Army investigators conducted experiments to determine the value, if any, of review of ratings. They found that the validity was improved a statistically significant amount and therefore included the review procedure in their new program.

The practice of requiring substantiating statements, (see page 34) though its primary purpose is to cause careful thought by the rater, may also be considered a measure to facilitate review. This is true even in a system in which ratings are accepted and filed as made. The substantiating statements will aid the central personnel agency in evaluating the ratings, attaching significance to those on which the substantiating statements indicate accurate ratings and giving less consideration to those on which the statements fail to support the marks given.

NEW APPROACH

The product of a relatively new approach to the problem of rating was placed in regular use by the Army in 1947. It employs the forced choice technique principally and, consequently, has come to be known as the Army's New Forced Choice Rating. It is the outcome of nearly two years work by a group of psychologists in the Army, during which they had the advantage of carefully controlled tests of the device in rating almost 50,000 officers. The results were reported to be quite good and the device was declared, "definitely

When preparing to implement a new rating device,[35] Army investigators conducted experiments to determine the value, if any, of review of ratings. They found that the validity was improved a statistically significant amount and therefore included the review procedure in their new program.

The practice of requiring substantiating statements, (see page 34) though its primary purpose is to cause careful thought by the rater, may also be considered a means to facilitate review. This is true even in a system in which ratings are accepted and filed as made. The substantiating statements will aid the central personnel agency in evaluating the ratings, attaching significance to those on which the substantiating statements logically accurate ratings and giving less consideration to those on which the statements fail to support the ratings.

THE REVIEWER

The product of a relatively new approach to the problem of rating was placed in reports use by the Army in 1947. It employs the forced choice technique, initially only compulsorily, but does so in such a way that "make-believe choice" doing. It is the method of testing by "pairs" by a group of psychologists in the Army. [illegible] have the advantage of objectively controlling [illegible] in rating almost 40,000 officers. [illegible] to be quite [illegible]

superior to any other yet devised and tested in fulfilling the requirements of an adequate rating system for Army purposes--..."[57]

The forced choice technique had been used with some success in personality measurement, The Jurgensen Classification Inventory and the Shipley Personal Inventory being notable examples. Their purpose, which they achieved to an acceptable degree, was to secure a true measure of an individual's personality despite the tendency of most people to conceal what they consider are undesirable or socially unacceptable aspects of their personalities. This technique for forcing the respondent to reveal his personality was adapted by the Army psychologists to forcing the rater to reveal his real opinion of the ratee.

A more detailed description and discussion of the forced choice rating will be found in Chapter 4, PART II.

NUMBER OF TRAITS

Mathematical considerations. Thus far the discussion has been concerned with attempts to improve reliability and validity by refinements of rating devices, design of different rating methods, and by various combinations of method and device. Mathematically it has been shown that increasing the length of a psychological measuring instrument increases the reliability and validity. The formulae showing these relations are as follows:

(a) The general form of the Spearman-Brown prophecy

superior to any other yet devised has proved in fulfilling the requirements of an adequate testing device for this purpose--..."[37]

The Forced choice technique has been used with some success in personality measurement. The Gordon (Identification Inventory and the Edwin Ghiselli Inventory could be valuable examples. Their purpose, which they achieve to an acceptable degree, was to focus the attention of an individual's personality despite the tendency of most people to choose what they consider are desirable or socially acceptable aspects of their personalities. This technique for testing, the opposite to reveal his personality was adopted by the army psychologists to forcing the saver to reveal his real opinion of the issue.

A more detailed description and discussion of the Forced choice testing will be found in chapter 4, THE TI-TLING OF TEST.

Methodical considerations. This for the Ghiselli has been concerned only without to improve reliability and validity of preliminaries of testing instruments, weight of different testing methods, and to vary the desirable methods of the view. Ghiselli's Inventory is the one by which increasing the length of a psychological scale, increased inter-rater reliability and validity. The Inventory's testing items were since are as follows:

(a) The general [illegible] of the [illegible]

formula for predicting the reliability of a test of increased length is:[58]

$$r_{nn} = \frac{nr}{1+(n-1)r}$$

In which:

r_{nn}	is the predicted coefficient of reliability of the lengthened test.
n	is the number of times by which the test has been lengthened.
r	is the coefficient of reliability of the test before lengthening.

(b) The general formula for predicting the validity of a test of increased length is:[59]

$$r_{(nx)y} = \frac{r_{xy}}{\sqrt{\frac{1 - r_{xx}}{n} + r_{xx}}}$$

In which:

$r_{(nx)y}$	is the predicted coefficient of validity of the lengthened test.
r_{xy}	is the validity coefficient of the test before lengthening.
r_{xx}	is the reliability coefficient of the test before lengthening.

A basic assumption underlying the above formulae is that each item of the lengthened instrument measures some relatively unique quality. Mathematically stated, the assumption is that the items correlate low with each other.

<u>In favor of a reduced number of traits.</u> There is a feeling that rating scale items generally do not meet this

Formula for prediction, the reliability of a test of increased length is:[52]

$$r_{nn} = \frac{nr}{1+(n-1)r}$$

In which:

r_{nn}	is the predicted coefficient of reliability of the lengthened test.
n	is the number of times by which the test has been lengthened.
r	is the coefficient of reliability of the test before lengthening.

(b) The general formula for predicting the validity of a test of increased length is:[53]

$$r_{(nx)y} = \frac{r_{xy}}{\sqrt{\frac{1}{n} + r_{xx} - r_{xx}}}$$

In which:

$r_{(nx)y}$	is the predicted coefficient of validity of the lengthened test.
r_{xy}	is the validity coefficient of the test before lengthening.
r_{xx}	is the reliability coefficient of the test before lengthening.

A fairly convenient method is the use of formula [illegible] that is the length of the lengthened instrument [illegible] reliability. [illegible] is [illegible] that is that [illegible] correlation [illegible].

[illegible]

[illegible]

requirement of uniqueness. Lawshe, in describing the halo effect, wrote that the tendency of raters to mark an individual about the same on each trait "suggests that the many different items are, in reality, measuring about the same thing..."[60] The seventy-seven raters participating in the study reported by Stockford and Bissell reacted to only three basic factors (technical proficiency, mental proficiency, and social proficiency) even though the rating scale they used was made up of considerably more than three items.[61] A factor analysis by Ewart, Seashore, and Tiffin of 1120 ratings made on a twelve trait scale caused the investigators to conclude that, "worker competency could be rated on one or possibly two 'traits' as well as it is now rated on the basis of twelve"[62]

<u>The case for a larger number.</u> An answer to such criticisms can be found in the writings of both Bingham and Probst. The former, in reply to the question of why raters should be required to record trait ratings when they correlate so closely with the final overall estimate, declared:[63]

> ...an overall judgement is more likely to be correct if made after the rater's attention has been focussed successively on several of the candidate's specific traits.

Probst has condemned what he calls "Overloaded Factors", items on rating scales which are too all-inclusive such as "quality of work".[64] Moreover, Ewart, Seashore, and Tiffin did not close the door completely on multitrait rating scales. After

requirement of antagonism. Hawkes, in translating the tale of Genji, wrote that the technique of relating to each individual about the same on each trait "suggests that the very different lives are, in reality, remarkably alike the same thing..."[60] The seventy-seven tales particularly in the study reported by Rosenberg and Sheldon seemed to only three basic factors (emotional sensitivity, conformity, and social practicality) even though the raters said they used was made up of considerably more than three traits.[61] A factor analysis by Sweet, Cameron, and Fitzpatrick (1960 reading) made on a twelve trait scale caused the investigators to conclude that, "worker competency could be traced to one or possibly two 'traits' as well as it is now traced to the basis of twelve."[62]

The case for a larger number. An answer to such criticisms can be found in (1) within a of both Jungian and Freudian. The former is apt to the question of why there should be required to record the a man's psychologically or diverbally with cultural universal meanings, Jaspers:[63]

> ...[illegible] Jungian is more likely to be answer if [illegible] a [illegible] has been [illegible] or [illegible] conditional [illegible] traits.

Freud's [illegible] called "[illegible]", [illegible] [illegible] [illegible]

reporting their findings as described above, they called attention to the specificity of those findings and advanced the hypothesis that improved ratings would be possible on a multitrait scale which was made up of a number of unique items.[65]

Looking at the field of physical science for a moment, one will recall that an accurate measurement in that sphere is not the result of a single observation, but rather, it is the average (or some other measure of central tendency) of the results of several observations. It is quite possible that the use of a rating scale which employs ratings on twelve or more items to measure "one or possibly two" factors of worker competence is a similar process of averaging the results of several observations.

The writer was unable to find reports of any actual reliability or validity comparisons between rating scales which had been shortened, because of the findings of a factor analysis, and the long forms of such scales. A comparison of that nature would be a fruitful research project, it is believed, and the writer submits that any judgement as to whether a short or a long rating form is the better should await the outcome of such a test.

A final consideration in favor of a multitrait scale emerges when one realizes that a rating on an abbreviated scale of one or two factors will afford little information on which either management or the employee can base action leading to improvement of the latter. For instance, one of

supporting social linkages as described above. I my belief extension to the specificity of those linkages and examined the hypothesis that improved welfare would be positively and multiplicatively related with the size of a number of unique items.[6]

Looking at the field of regional science for a moment, one will recall that an analytic framework in this sphere is not the result of a single development, one reason, is the sheer size (or some other characteristic) of the region (or some other characteristic) of the region of general concern. It is quite possible that the use of a model which would employ matrices of twelve or more items to measure "how or whether" interaction of various components in a dynamic process of employing the results of previous interactions.

The relative ease with which the reports of any actual relationship or mutual dependence between matrix elements which are real anywhere, regardless of the size of a function or a figure, and i.e. for, items of such a kind. A consideration of this issue would be a limited research project. It is believed, that the present analysis does say [illegible] to whether a set of a few matrix elements is an best [illegible] measure [illegible] site.

the factors isolated by Ewart, Seashore, and Tiffin was "Ability to do Present Job". Is it likely that management or the ratee can do much toward improving the ratee in this factor unless his specific shortcomings as related to job performance are known?

The importance of the rating in employee improvement was stressed early in this thesis (pages 2 and 3) and it will receive further attention later in connection with the treatment of the practice of showing the ratee his ratings and discussing them with him. No design feature which seriously limits the usefulness of the rating in the important function of employee improvement should be adopted.

Practical considerations. Should the rating scale designer resist the blandishments of the advocates of the abbreviated scale and plan a many-itemed device to aid the thinking of the rater, as suggested by Dingman and by Probst, and to provide a basis for employee improvement, he must, nevertheless, guard against making his scale too long. He must give due consideration to the very real factors of cost of the form and of its administration, and he must have care lest he induce a perfunctory attitude in the raters by foisting upon them a form which requires an excessive amount of their time and energy.

ADMINISTRATION OF A RATING PROGRAM

Showing ratings to ratees. Implicit in the objectives of rating is the requirement that the ratee see his rating.

the factors included by Lewis, Gantt, and Hillis in "Profit-ly to do research job". It is likely that management or the research is much towards improving the factor which is specific characteristics related to job performance are absent.

The importance of the meeting in employee improvement was stressed early in other studies (Lewis & and 3) and it will receive further attention later in connection with the management of the practice of showing the value its results and changing them with them. No design feature which materially lim-its the usefulness of the meeting in the improvement function of employee improvement should be adopted.

Practical considerations. Finally the meeting must be adapted under which the dimensions of the [illegible] of the above-viewed media and plan a [illegible] device to aid the training of the raters, as suggested by Figure [illegible] and to provide a basis for employee improvement. In most, [illegible] [illegible], [illegible] meeting [illegible] [illegible] [illegible] [illegible] due consideration to the very real factors of cost of the form and of its administration, and to most [illegible] [illegible] be indeed a [illegible] attitude in the [illegible] by [illegible] upon [illegible] [illegible] [illegible] [illegible]

Halsey,[66] Yoder,[67] and Probst,[68] in common with many other authorities in this field, have agreed that provision should be made in every rating program for the employees to see, or be informed of, their ratings. Scott, Clothier, Mathewson, and Spriegel also recommended a post-rating interview between rater and ratee and proposed the following statements for the rater to complete in planning the interview:[69]

How I Can Help Him To Be More Effective
On His Present Job:

He should be given additional instruction on _____
He should be given additional experience on
such jobs as _____
He should study such subjects as _____
He should change his attitude as follows _____
There is nothing more I can do for him because _____
Remarks: _____

More recently, Armstrong wrote of the post-rating interview.[70] After some discussion, he proposed the form shown in Figure 2, page 54, to be printed on the rating blank and to be used in preparing for the interview. In addition, he offered the suggestions for actual conduct of the interview shown in Figure 3, page 55.

<u>Who should rate.</u> Obviously he who has the most accurate knowledge of the ratee should perform the rating and no disagreement relative to the statement as voiced here will be encountered. However, the question often arises as to whether it might not be better to utilize ratings made by the ratee's peers or his subordinates instead of those performed by his superiors.

Haines,[86] Yoder,[87] and Probst,[88] in common with many other authorities in this field, were agreed that provision should be made in every radio program for the employees to see, or be informed of, their ratings. Scott, Clothier, Spriegel, and Dyrenforth also recommend a centralized interview between rater and ratee and proposed the following steps for the rater to complete in learning the interview:[89]

How I can help this to be more effective
In this work for:

He should be given additional instruction in
He should be given additional experience on
work in
He should study more thoroughly on
He should be given a job original or follows
There is nothing more I can do for the person
Therefore:

Type Personnel, Training [illegible] a motivated [illegible] terviews.[90] [illegible] in Figure 2, [illegible] to be used in [illegible] offered [illegible] shown in [illegible].

[illegible] rate, [illegible] discussed [illegible] and [illegible] in [illegible] per [illegible] way [illegible].

RATING FORM

TRAITS :_____:

:_____:

:_____:

PLAN YOUR TALK
TO MAN

TALK TO HIM
ABOUT THESE
FIRST

THEN TALK TO HIM
ABOUT THESE

(Favorable Remarks)

(Unfavorable Remarks)

FIGURE 2

FORM TO AID LEADER
IN PREPARING
FOR POST RATING INTERVIEW

STEP I	STEP II	STEP III
Start out by stressing good points of your rating.	Follow with what's wrong with the record.	Repeat and conclude with other good points in the man's record.
→ Talk → Good worker.	→ This → Attendance poor.	→ Way → Knows his job.

FIGURE 3

PROCEDURE FOR [illegible] RATING INTERVIEW

There is a general feeling that ratings by one's peers, the so called "buddy ratings", are more accurate than supervisors' ratings. Army psychologists used "buddy ratings" to standardize the Army Forced Choice rating form[71] and subsequently, Wherry and Fryer reported finding clearcut superiority in "buddy ratings".[72]

An Army study reported by Bittner showed that 78% of a group of 1800 Army officers preferred not to be rated by officers of equivalent rank while 77% of these same 1800 officers preferred not to be rated by their juniors.[73] If these attitudes are typical, it seems likely that rating by one's superiors will continue as the usual practice and there remains only the problem of selecting as the rater, the superior with the most adequate knowledge of the ratee.

<u>How often to rate.</u> Although there has been no experimental evidence obtained to answer this question, authorities writing in the field are generally in accord in their recommendations. Halsey[74] and Probst,[75] for example, recommended ratings twice each year, as routine, with more frequent ratings in special situations such as a rapidly expanding or shrinking organization and new or probationary employees. Probst suggested in addition, that ratings be performed on employees about to be transferred. Both men oppose the practice of allowing an excessive period to elapse between ratings, warning of the danger that the raters' memories will be unequal to the test which will lead them to give undue weight

There is a general feeling that rating by one's peers, the so called "buddy rating", are more accurate than superiors' rating. Army psychologists used "peer rating" to standardize the Army Ground Combat Rating Form[71] and subsequently, Storry and Pyper reported finding considerable superiority in "buddy rating".[72]

An Army study reported by Litzner showed that 71% of a group of 1000 Army officers preferred not to be rated by officers of equivalent rank; 71% of these same 1000 officers preferred not to be rated by their juniors.[73] If these attitudes are typical, it seems likely that rating by one's superior will continue as the usual practice, but there remains only the question of reliability of the rater, the supervisor with [illegible] responsibility in the process.

The officer to rater. Although there has been considerable evidence [illegible] to answer this question, [illegible] written in [illegible] general. [illegible] mentioned. [illegible][74] and [illegible][75] or [illegible], [illegible] rating [illegible] [illegible] [illegible]

to the more recent performance of the ratees.

Rating the rater. It has been suggested by some that poor raters be identified and relieved of the responsibility. A capacity for correctly judging subordinates is certainly a desirable attribute of one in authority, and in the light of that consideration, the performance of a rater may be seen as one measure of his success as a supervisor. Further, ratings of raters would be an aid in rater training just as ratings of employees are helpful in employee training.

Probst recognized the need for locating the incompetent raters and pointed out how his rating system will facilitate this.[76] Yoder discussed "rating the raters" in connection with his plan for correcting ratings to a comparable basis[77] and White proposed statistical analysis of ratings in order that, "... inexpert or careless rating officials can be confronted with the sometimes foolish consequences of their lack of care."[78]

TRAINING OF RATERS

The foregoing discussion has been concerned largely with technical improvements and design of rating devices and procedures as they relate to the struggle for increased reliability and validity. Valuable as these elements are, most authorities agree that the most significant factor in a successful rating plan is an adequate rater training program. Turning to the writings of White one finds the following passage:[79]

to the more recent performance of the reviews.

Rating the raters. It has been suggested by some that poor raters be identified and relieved of the responsibility. A capacity for correctly judging subordinates is certainly a desirable attribute of one in authority, and in the light of that consideration, the performance of a rater may be used as one measure of his success as a supervisor. Perhaps, ratings of raters could be on file in rater evaluation (as) ratings of employees are used in employee evaluation.

Proper management use need for improving the interviewing raters and helping them develop more effective rating techniques tools.[76] John Hickman "reviews the rater" in connection with the planning conference with the subordinate. "... Inquiry or conclusions which officials can be confronted with the questions which subordinates of their lack of case."[77]

Theory and Practice

The foregoing discussion has been concerned largely with conceptual improvements and tools or tools, judging and procedures as well rather than a discussion for quality of facilities and validity. [illegible] subordinate [illegible] essential [illegible] [illegible] [illegible]

> It is indeed now generally held that the form of the rating instrument is less important than thorough training of the rating officers in the art of evaluating subordinates.

Bittner,[80] Knowles,[81] and Tiffin,[82] to mention only a few others, voiced similar views, while Driver[83] even went so far as to advocate the retraining of raters at intervals.

Effect of training measured. In 1945 the research section of the Army Adjutant General's Office conducted a carefully controlled experiment to determine the results of rater training. The population of officers, who were to act as raters, was separated into higher echelon and lower echelon categories and organized into control and experimental groups. After all subjects had performed sample ratings (which were correlated against certain criteria to determine validity), the experimental groups were given a two hour course of instruction and practice in rating. Upon completion of this training, all the members of the control and experimental groups rerated the individuals they had evaluated in the preliminary operation. Correlations between rating validities showed a slight and consistent, though not statistically significant, gain for the higher echelon officers but showed no consistent improvement for the lower echelon officers.[84]

The experimenters commented, when discussing the small improvements obtained, that their subjects were all experienced raters who could not be classified as untrained, and conse-

It is indeed now generally held that the form of the radio instrument is less important than thorough training of the rating officers in the art of evaluating subordinates.

Dittner,[80] Moyles,[81] and Griffin,[82] so constitution only a few others, voiced similar ideas. Mr. Taylor[83] even went so far as to advocate the retention of rater by interviews.

Effect of training upon ratings. In 1919 the Research section of the Army Distinct Personnel Office conducted a carefully controlled experiment to determine the results of rater training. The population of officers, who were to act as raters, was separated into lower section and lower section officers and organized into control and experimental groups. After all subjects had performed certain rating (which were correlated against other criteria to obtain the validity), the experimental group was given a [illegible] course of instruction and practice [illegible]. Upon tabulation of this training, all the [illegible] control [illegible] mental group showed a [illegible] in the predictive operation. Dorsey[84] in Lorenzen [illegible] Militiae showed a [illegible] by [illegible], when [illegible] certain [illegible], [illegible] showed [illegible] lower [illegible] officers.

quently, the effect of training could not be expected to appear as marked as if administered to untrained raters. To the writer it seems that the relatively small effect of training experienced raters seen in this study would tend to weaken Driver's recommendation that provisions be made for retraining raters at intervals.[85]

Effect of certain rater qualities. A more recent study by Stockford and Bissell revealed that training in rating reduced the susceptibility of raters to bias and to halo and increased the reliability of their ratings, although it had little effect on the rater's leniency. Other interesting data were obtained when several qualities of the raters were measured by standard psychological tests and correlated with various aspects of rating performance. The higher the rater's mental maturity, the more reliable, less lenient, and less biased will be his ratings. Moreover, as one might expect, the more intelligent raters profit more from the rater training. Lastly, the higher the rater's Persuasive Component and his Administrative-Supervisory Interest factor, as determined by the Kuder Preference Record, the less lenient and biased will be the ratings made by him.[86]

It appears then that the logical assumption that rater training is a valuable adjunct to a rating program is borne out by test. Therefore consideration of what material should be included in a rater training program is in order.

quently, the effect of training could not be expected to support as marked as it obtained in previous studies. To the writer it seems that the relatively small effect of training experienced raters may in some sense be traced to weaker Drivers's instructions than provided in such for rerating rater as interests.[28]

Effect of certain rater qualities. A more recent study by Thorndike and Hagen revealed that training in rating reduced the probability of rater's in line and to halo and improved the reliability of their ratings, although it had little effect on the rater's leniency. Other investigating have been obtained when several qualities of the raters were considered, several studies have reported [illegible] with various [illegible] [illegible]. [illegible] raters's mental ability, [illegible] [illegible] leniency will [illegible]. [illegible] [illegible] [illegible]

Substance of Rater Training Program

The decision as to what should be taught rests on the determination of what are the most common rater errors, since it is toward the correction of these that the program should be aimed. Because the most frequently occurring raters' errors have been enumerated and discussed at some length in the preceding chapter, they will not be detailed here. Instead the discussion will be limited to the training program proper.

The purpose of rating. The most suitable starting point in the rater training program is emphasis on the role of the rating device as a measuring instrument. In this connection, the objectives of rating, as discussed on pages 3 and 4, should be made clear to the raters. The concept of the rating as a means of personal reward or punishment should be eliminated. The value to the Navy, and the Nation, which will accrue from a well designed and properly used rating system should be stressed. Conversely, the harm to the Navy, and the Nation, which will result from a rating system which has been invalidated by its use as a reward or incentive should be impressed upon the raters.

Theory of normal distribution. The theory of the normal distribution and its relation to individual differences should be explained. Examples of actual studies which demonstrate its truth should be cited. The absurdity of the feeling that Naval officers must be "above average" to be of value to the Naval Service should be shown. If these points are

Substance of motor training program

The decision as to what should be taught rests on the determination of what are the most common mistakes, since it is these the correction of which can be expected should be aimed. Because the most frequent, wrecking mistakes have been enumerated and discussed at some length in the preceding chapter, this will not be needless work. Instead the discussion will be limited to the training program, [illegible]

The purpose of training. The most vital in shooting points in the basic training, [illegible] the testing device as a [illegible] instrument. In basic commission, the objectives of training, [illegible] should be made clear to the trainee. The [illegible] ing as a means of personal mastery [illegible] attained. The value to the [illegible] will appear [illegible] system should be [illegible] and the [illegible] has been [illegible] should [illegible]

clearly and convincingly made, significant progress towards enhanced reliability and validity will have been achieved.

Psychological influences. The more important psychological phenomena which influence the raters should be identified, described, and discussed. The raters should be informed of the effects of these forces and warned of their subtle nature in order that they may be on guard to minimize their consequences.

Meaning of terms. Explanation and discussion of the meanings of the terms used on the rating form are important. The instructor should make frequent use of examples and anecdotes to clarify his explanations. Having the student raters take turns at composing descriptions of fictitious, or actual situations which illustrate the meanings of the terms will be especially beneficial. Successful teaching of this section will do much to eliminate the difficulties of differing standards of comparison as well as minimizing errors due to misunderstanding of terms.

Practice ratings. Learning by doing is conceded to be the most effective type of learning when the amount of learned material retained and the ability to apply this learning are the criteria. Consequently, no rater training program would be complete without practice rating. A suggested procedure is to require the student raters actually to perform ratings on real persons or on the basis of anecdotal records. These anecdotal records may be of real or imaginary persons but

clearly and conveniently made, significant progress towards enhanced reliability and validity will have been achieved.

Psychological influences. The more important psychological phenomena which influence the raters should be identified, described, and discussed. The raters should be informed of the effects of these forces and warned of their subtle nature in order that they may be on guard to minimize their consequences.

Meaning of terms. Explicit and discussion of the meaning of the terms used on the rating form are important. The instructor should make frequent use of examples and anecdotes to clarify his explanations. Having the student masters take turns at composing descriptions of fictitious, or actual situations which illustrate the meaning of the terms will be especially beneficial. Successful teaching of this section will go much to minimize the difficulties of rating standards of comparison as well as eliminating errors due to misunderstanding of terms.

Practice ratings. Learning by doing is conceded to be the most effective type of learning. Since the amount of learned material retained and the ability to apply this learning are the criteria, necessarily, no real training program should be complete without practice ratings. A few general guidelines are in regard to practice ratings. These actually are rating on real persons, or on the basis of film or tape recordings. These should be made [illegible]

should be so selected that they bring out the many fine points of the rating problem.

Comparisons and correlations of the ratings on the same persons would show their reliability or lack thereof. Subsequent discussion and further practice would bring improved rating performance.

Practice rating would have the added beneficial effects of:

a. Familiarizing the raters with the forms and procedures.

b. Standardization of the meanings of the terms used on the forms.

c. Elimination of or reduction of differences in standards of comparison among the raters.

SUMMARY OF BEST IN RATING DEVICES AND PROCEDURES

In the preceding sections of this chapter an effort has been made to describe and evaluate the most common types of rating devices and procedures together with descriptions of the more important relevant expert opinion and experimental evidence. From this study there emerges a body of principles of good service rating design and administration. The remainder of the chapter will be devoted to the compilation of this information.

Principles Applicable to Devices

<u>Bias and objectivity.</u> The rating device should embody features which aid in minimizing the rater's bias and improving

should be so adjusted that they bring out the many fine points of the rating problem.

Comparisons and correlations of the ratings on the same persons would also throw reliability or lack thereof. Moreover discussion and further practice would develop improved rating performance.

Possible rating, would have the other beneficial effects of:

a. Familiarizing the raters with the forms and procedures.

b. Standardization of the meaning of the terms used on the forms.

c. Elimination of or reduction of differences in standards of comparison among the raters.

CONCLUSIONS

In the preceding sections of this chapter an effort has been made to describe and evaluate the current form of rating ratings and procedures employed by the Naval Ships. More important relevant aspects have been emphasized. From this study, there seems to be a need for improvement in the rating forms and procedures and standardization. The results of the study indicate a desire on the completion of the rating forms, rating procedure applicable to officers.

Bias and subjectivity. The rating forms are subject to several features which tend to influence the ratings in and improper ways.

his objectivity. Further, it must eliminate insofar as possible, errors due to misunderstanding of terms and to differences among raters in standards of comparison. These objectives are best served by the forced choice type of rating. However, if the expense of preparation and maintenance of the forced choice form is prohibitive, or if it meets objection for some other reason, both the Probst check list type of form and the graphic rating scale are possibilities. If the graphic scale is chosen, it should be of the descriptive rather than the evaluational type and consideration should be given to the possibility of requiring substantiating statements with the ratings. The various other rating processes, such as the ranking and the paired comparison methods, will be eliminated from further discussion except as one element of a combination form, to be discussed in a moment.

Halo. The construction of the rating device should provide characteristics which will aid the rater to avoid the halo error. Here again the forced choice form is preeminent. Other possibilities are the Probst scale and the graphic scale with the favorable ends of successive trait scales alternated irregularly from right to left and back again. Other features include unequal graduations of successive trait scales or elimination of graduations altogether, leaving only the labels.

Normal distribution. An attempt should be made to secure a reasonably close approximation of a normal distribu-

bia objectivity. Further, it must eliminate insofar as possible, errors due to measurement or to factors not to differences among raters in standards of comparison. These objectives are best served by the forced choice type of rating. However, if the expense of preparation and maintenance of the forced choice form is prohibitive, or if it seems objection for some other reason, then the forced check list type of form and the graphic rating scale are satisfactory. If the graphic scale is chosen, it should be of the descriptive rather than the evaluational type. The information should be given to the possibility of receiving accomplishing statements with the rating. The various other rating procedures, such as the ranking and the paired comparison methods, will be eliminated from further discussion except in our elements of a combination form, be of limited use in a manual.

Info. The use (type) of the rating data is to provide a measurement which will be taken to avoid the main error. Here again the forced choice form is superior. Other possibilities are the check scale and the graphic scale with the descriptive mode of measurement. Other scales listed from which to left and back scale elements are listed. Can which to left and back scale. Other factors include respect variations of measure- ive trait scales or elimination of relationship biases, halo, and the halo.

Forced evaluations. A rating form is to be se- sure - evaluations, - - - - - elimination of - - - - - -

tion of rating scores. With the exception of the forced distribution rating (which has been declared unsuitable for Navy use), the Probst scale seems to be the most successful at attaining this objective with the forced choice rating as second best. Improvement of the graphic form in this respect is attempted by dividing its scales into columns, as described earlier, and supplying hortatory instructions that the raters be guided by the columns when rating.

Statistical methods of correcting ratings may be employed. The standard score (z score) technique previously discussed will accurately convert rating scores to comparable quantities on a normal distribution. Alternative methods, one a mathematical process using average deviations rather than standard deviations, and the other a graphic process, are suggested by Yoder.[88]

Weighting of traits. There should be provision of suitable weighting of traits in arriving at the overall score and precautions should be set up against uncontrolled weighting. Such precautions are best provided by the standard score (z score) method or some approximation of it.

Display of qualities. The rating should be in such a form that both the ratee and management can see what are his strong points and his weaknesses. If this condition prevails, both the ratee and management can take intelligent steps to correct the weaknesses. In addition, management may wish to institute procedures which will lead to selection of superior

[illegible]

[illegible]

[illegible]

personnel in the future.

This requirement that there be an intelligible display of the rater's qualities is best met by the graphic rating scale, though the Probst scale is a possibility. The forced choice form is poorest in this respect since by its very design the true nature of the rating is concealed from rater and ratee.

Combinations. It is common with military rating forms to include two or more methods of rating in the same rating instrument. On the basis of custom and logic, experimental evidence pro or con not being available, the practice should be continued.

Traits to be excluded. The rating scale's raison d'être is the fact that certain qualities of men are measurable only by judgement. If a better method of measuring these certain qualities existed it would be only wise to dispense with rating scales and use the better method. Conversely, it is wise measurement policy to exclude from rating scales those qualities which can be measured better by more objective means. Included in this category are such entities as "health", "intelligence", "accuracy", "speed", "volume of production", and "attendance". While "personality" has defeated all attempts at accurate objective measurement, it is too large and all-inclusive a quality to be entrusted to the limited capabilities of rating scales.

personnel in the library.

While respondents may stress the intellectual liveliness of the writer's qualities in such matters as reading and teaching, through the broad scale is a possibility. The forced choice form is, however, in this respect since by its very design the true nature of the writing is concealed from view and nature.

Conclusions. In the common [illegible] library rating forms to include two or more members of staff in the same rating instrument. In the [illegible] of [illegible], supplemental evidence pro or con [illegible] available, the practice should be continued.

Trails to be avoided. The rating scale's raison d'être is the fact that certain qualities of man are measurable only in [illegible]. [illegible] these certain qualities subject to rating only also to disperse with rating scales [illegible] in the [illegible] above. Conversely, it is also [illegible] to conclude from [illegible] these qualities which [illegible] by more objective means. Included in [illegible] qualities is "[illegible]", "intelligence", "[illegible]", "[illegible]", "values in [illegible]" and "[illegible]". [illegible] "[illegible]" in [illegible] as accuracy of [illegible] measurement. It is [illegible] [illegible] directly [illegible] interpreted to [illegible] qualities [illegible] rating scales.

Administrability. Rating instruments and procedures should not require excessive time and energy to employ and administer. This requires that the forms be of a convenient size with easily readable printing, and that they not require an excessive number of operations. Scoring aids such as templates and stencils, or better still, machine scoring, should be provided. Finally, the data produced by the device should be readily obtainable from it in an easily used form.

Principles Applicable to Procedure

Conference with ratee. Definite provision for discussing the rating with the ratee should be included in every rating program. The ratee should be shown his detailed rating, complimented on his strong points, and sympathetically and constructively counselled on means of overcoming his weaknesses.

Who should rate. The consensus is that the superior who has the most intimate knowledge of the ratee is the man who should rate him. Where more than one superior is available with adequate knowledge, the rating procedure should be designed to take full advantage of the superiority of several judges over one. Either a conference rating or the average of several individual ratings may be used, but the latter is the more accurate.

How often to rate. Informed opinion suggests that semiannual ratings are sufficiently frequent except when the organization is expanding or contracting rapidly, at which time a shorter period, usually three months, between ratings

<u>Administerability.</u> Tests, instruments and procedures should not require excessive time and energy to employ and administer. This requires that it be done by a convenient size with easily readable printing, and that they not require an excessive number of operations. Scoring aids such as templates and stencils, or better still, machine scoring, should be provided. Finally, the cost produced by the device should be readily obtainable from it is so easily used from.

Principles applicable to procedure

<u>Conference with pupils.</u> Definite preparation for discussing the testing with the pupils should be included in every testing program. The pupils should be fully acquainted with complications on the answer points, and improvementally and constructively treated as means of preparing is examinees.

<u>The should pupil.</u> The counselor is not the superior who has the so i intimate knowledge of the pupil in the use who should note the. [illegible] the counselor is available with adequate knowledge, the pupil [illegible] procedure should be designed to take full advantage of the superiority of personal judgment over one. Either a conference within or the average of several individual [illegible] is best, but the lesson is the more accurate.

<u>The pupil to note.</u> Informed children usually [illegible] conditional [illegible] sufficiently [illegible] adopt [illegible] organization [illegible] or motivation [illegible], [illegible] [illegible], [illegible] [illegible], [illegible]

is recommended. In addition, it is believed that new members of the organization should be rated more frequently, because, (1) they are likely to be changing more rapidly as they learn their duties than are the experienced employees and, (2) more frequent ratings will provide a sounder basis of evaluation during and at the end of the probationary period.

Review of ratings. Where practicable, advantage should be taken of the tendency of raters to give more thought and care to their rating tasks when they know that they may have to justify their judgements to their superiors. This is best done by some procedure for formal review and endorsement of ratings.

Training of raters. A rater training program should be maintained which not only acquaints the rater with the mechanics of the use of the form, but impart some understanding of the more important theories underlying rating.

Rating raters. Just as employee rating is a tool for employee improvement, so ratings of the raters (by the authorities in the central personnel agency and with adequate liasion with the rater training program) will be a useful tool in improvement of the raters.

is recommended. [illegible]

of the [illegible]

(1) [illegible]

their [illegible]

frequent [illegible]

during [illegible]

[illegible]

should be [illegible]

and [illegible]

have [illegible]

is [illegible]

ment [illegible]

[illegible]

be [illegible]

[illegible]

[illegible]

[illegible]

equipment [illegible]

thus [illegible]

lizer [illegible]

too [illegible]

PART II

Chapter 3

CRITERIA

Importance of Criteria

Measurements are merely comparisons. It is impossible to conceive of a measurement that is not a comparison for such would be meaningless. In the physical sciences comparisons can be made between the physical entity being measured and a measuring instrument, which in its design and construction has already been compared with a standard, which in turn has been compared with a more refined standard and so on, back to the fundamental standard, the criterion. Measurement in the social sciences requires standards also. Here too, one must have something with which to compare his measuring instrument (a criterion) in order to determine the accuracy of the instrument.

Inadequacy of Available Criteria

In general, criteria available to social scientists have not been as reliable and valid as those at the disposal of the physical scientists. This has been particularly true in the field of measurement of executive ability, which is the area in which this study of officer efficiency evaluation is concerned.

<u>Criteria of reliability.</u> Normally, the reliability coefficient of a rating device will be determined by correlating

PART II

Chapter 6

UNIT IX

Importance of Units

Measurements are usually inaccurate. It is impossible to conceive of a measurement that is not approximate; for such would be meaningless. In the physical sciences experiments can be made because the physical world only measured and a measuring instrument, which is for design and construction has already been constructed with a [illegible] which is found has been complete, also [illegible] relative [illegible] so on, back to the fundamental standard, the arbitrary. Experiment in the social sciences involves another - also. Here too, one must have a standard with which to compare the measurement instrument (a questionnaire) in order to determine the accuracy of the instrument.

Importance of Reliability/Validity

In general, scientific [illegible] it would seem should have not been so [illegible] a [illegible] of the physical sciences. [illegible] in [illegible] the [illegible] to [illegible]

[illegible] of Reliability

[illegible]

the results of one of the following alternatives:

Alternative	Ratings made by.	Number of ratings.	Forms used.
1	Both by the same rater in each individual case.	Two	Two different but comparable.
2	By two different raters.	Two	Same form for both ratings.
3	Same raters.	Two spaced some time apart.	Same.

Typical coefficients of reliability. Yoder has set $r = +.75$ as the minimum reliability coefficient which should be tolerated.[1] By performing the following calculation of forecasting efficiency:[2]

$$100(1-\sqrt{1 - r^2}) = 100(1 - \sqrt{1 - (.75)^2}) = 33.9\%$$

we see that Yoder's limit requires that the rating device be reliable only to the extent that the likelihood of two consecutive ratings of the same individual being identical is 33.9% better than chance. That is to say, Yoder's limit requires that the reliability be such that one has 33.9 chances in one hundred of predicting what a second rating would be. Although rating designers have enjoyed some success in exceeding this minimum, none has seen fit to propose raising it.

Probst reported finding a reliability, in one case of ratings of 475 employees, of +.72 and added that this was typical of the experience with his device.[3] Other investigators, to mention only a few, have reported reliability coefficients from their studies as follows:

the results of [illegible] following administrations:

Alternative	[illegible]	[illegible]	[illegible] of Alternatives	Power curve.
I	[illegible]	[illegible]	Two	Two different [illegible]
II	[illegible]	[illegible]	Two	Two [illegible] for [illegible]
III	[illegible]	[illegible]	Two groups [illegible]	[illegible]

Tables constructed of reliability. Later has not $r = +.71$ as [illegible] reliability coefficient which [illegible] be balanced. By comparing the following calculation of frequency differences:

$$100(1 - [illegible]) = 100(1 - \sqrt{1 - [illegible]}) = 50.0$$

we see [illegible] reliable [illegible] outline [illegible] 53.9% [illegible] quite [illegible] in [illegible] Alternative [illegible] the [illegible] [illegible] [illegible]

Stockford and Bissell[4]	+.798
Richardson and Fuder[5]	+.85 to +.90 (N = 330) and +.83 (N = 305)
Lawshe, Kephart, and McCormick[6]	+.83

One of the highest coefficients of reliability for a service rating likely to be found is the r = +.92 reported by Tiffin.[7] It is significant that the system which yielded this high figure employed "systematic corrections" to equalize "rater tendencies" and independent ratings of each ratee by several raters. Only the designers of the U. S. Air Force officer rating system, see Chapter 4 of PART II, report a higher coefficient.

Criteria of validity. The followin passage from a recent book by Lawshe describes the situation in this area succinctly and well:[8]

> ...A Survey of the literature yields little evidence of successful validity studies in the executive brackets. This is no doubt due in part to the extreme difficulty attending the setting up of adequate criterion groups at the executive level, ...

This problem of absence of criteria arises from the very condition which makes it necessary to depend on mere judgement in this important field of measurement, i.e., from lack of a better method of measuring many of the aspects of executive performance. To secure a criterion, one must in some way, identify and isolate a standard of executive performance. He must select a definite sample of executive per-

Stoolfard and Hassell[4]	+.703
Riverdale and Yuba[5]	+.60 to +.90 (N = 380) and
	+.05 (N = 305)
Lawson, Lapham, and Schneider	.00

One of the highest coefficients of reliability for a service rating blank, [illegible] r = .92 reported by Griffin.[7] It is significant that [illegible] higher than high figure employed "systematic correlations" to eliminate "halo tendencies" [illegible] of each rater by several raters. [illegible] of the [illegible] service officer rating system, the [illegible] r = .71 indicates a higher coefficient.

Criteria of validity. The following passage from a recent book by Lawson describes the situation in this area succinctly and well:[8]

> ... survey of the literature yields a little evidence of a consistent relation between the executive variables. This is no doubt due in part to the extreme difficulty attending the setting up of adequate criterion measures of the executive [illegible].

This [illegible] of criteria makes for the very difficult [illegible] in research in [illegible] Judgment is the important factor of [illegible] lack of a better [illegible] the [illegible] of the executive [illegible] executive performance. [illegible] ability, [illegible] [illegible]

formance, that is, measure a finite quantity of it. Except when measuring certain qualities such as intelligence, aptitudes, and health, judgement is the only recourse in accomplishing this appraisement. The result is comparable to the outcome of an attempt to test the accuracy as to length of a "one foot" ruler which you have made, by comparing it with a friend's estimate of how long "one foot" is.

The frustrating condition described in the preceding paragraph has led Cronbach to the hopeful remark quoted earlier that, "Observations by impartial observers are generally accepted as valid if they can be made reliable."[9] Too often this is exactly the situation which prevails. But, in the writer's opinion, it is an unsound position to adopt. It is quite true that validity depends on reliability, that there can be no validity unless there is reliability. The mathematical statement of this relation is shown on page 49. But it is not true that because the measuring instrument is reliable, it is therefore a valid measure <u>for the particular quality</u> being considered at the moment. Again we may look to the realm of physical measurement for an analogy. If a reliable measure is also necessarily valid, then a ruler will yield valid linear measurements even if through some inadvertence it had been made actually too short. Obviously, this cannot be, for though readings obtained with it were uniform, they would be nonetheless incorrect.

[illegible] when measuring certain qualities such as intelligence, aptitudes, and health, judgment is the only resource in accomplishing this appreciation. The result is comparable to the outcome of an attempt to test the accuracy of the length of a "one foot" ruler which [illegible], by comparing it with a friend's estimate of how long "one foot" is.

The [illegible] conclusion included in the preceding paragraph has been reduced in the [illegible] [illegible] quoted earlier that, "[illegible] of [illegible] observers are [illegible] accepted as valid if they [illegible] reliable."[9] Too often this is exactly the [illegible] [illegible] reality. [illegible], in the writer's opinion, it is an [illegible] position to adopt. It is quite true that validity depends on reliability, [illegible] can be no validity unless there is reliability. The [illegible] mathematical statement of this relation is [illegible]. But it is not true that [illegible] [illegible] reliable. It is therefore [illegible] <u>for the particular quality</u> being [illegible] [illegible] to the test of [illegible] [illegible]. If a reliable measure is [illegible] [illegible] will yield valid [illegible] [illegible] tance it has been [illegible] [illegible]. Obviously, this cannot be, for there is [illegible] [illegible] [illegible] [illegible].

Typical criteria of validity. Many investigators in this field have relied simply on face validity and reliability, as described under the graphic rating scale in the preceding chapter. Obviously no coefficient of validity is available from such studies since the reliability coefficient cannot be substituted for it. Those experimenters who have attempted validation proceedings have utilized, in most cases, some other form of rating as the criterion. A "validity coefficient" of a sort is produced by this process but its exact value may be open to some doubt.

Numerous examples of the validation of one rating by the use of another as a criterion may be found in the literature of rating. Consider Probst's recent book on the subject.[10] He described his criteria as "supervisors' judgements" and reported correlations between his rating device and these as validity coefficients. A few of these are shown in the following table.

Page	Validity	Number of cases	Criterion
78	+.66	25	Ratings by superintendent.
78	+.84	32	Average of judgements of 3 supervisors.
79	+.77	22	Average of judgements of department head and immediate supervisor.

<u>Typical criteria of validity.</u> Many investigators in this field have relied simply on face validity and reliability, as described under the rating scale in the preceding chapter. Advocates of coefficients of validity in excess from zero at once raise the reliability coefficient cannot be established for it. Those objectivists who have attempted validation procedures have utilized, in most cases, some other form of rating as the criterion. A "validity coefficient" of a rater is measured by how closely his ratings agree with those of other raters.

Interesting examples of the validation of one rating by the use of another in a criterion may be found in the literature of rating. Cunningham [illegible] research upon the subject.[10] He described his subjects as "experienced judgments" and reported correlations between his rating scales and there is reliable coefficients. Two of these are shown in the following table.

Page	Validity	Number of Cases	Criterion
72	.08	60	Rating [illegible]
70	[illegible]	70	[illegible]
72	[illegible]	70	[illegible]

Another and less extensive study of validity using supervisors' judgements as criteria was reported by Knauft.[11] He obtained validity coefficients for two separate rating devices of +.87 and +.83 respectively.

Ratings by peers as criteria. The search for validity criteria has led to use of the sociometric method.[12] This is the "buddy rating" procedure or the "nominating technique" so often mentioned in current writings on merit rating. It is a process in which the ratees rate each other anonymously and, as used by rating scale designers, may result in orders of rank or a complete rating of each ratee on a list of traits. Correlation between the result of ratings on the new rating device and the results of the buddy rating will produce a validity coefficient.

Ratings by peers have been stoutly defended by Sherry and Fryer.[13] Dr. Sherry is one of the psychologists who participated actively in the work of design, construction, and standardization of the Army forces officer rating device, and it was in this work that buddy ratings were extensively used.[14]

The study reported by Sherry and Fryer [15] was made at Army Signal Corps Officer Candidate School at Fort Monmouth, New Jersey in 1943 and involved two separate classes, one containing eighty-two men, the other only fifty-two. Both buddy rankings (nominations) and buddy rating data were obtained and compared with ratings by junior officers and senior officers on the school staff. Correlations were also obtained between the

[illegible]

[illegible] supervisory [illegible] [illegible]

He obtained satisfactory qualitative [illegible] [illegible]

vision of 1.5V and less respectively.

Methods of [illegible] the [illegible]

criteria [illegible] [illegible] criteria.[illegible] While in the "hard" working [illegible] "[illegible]" [illegible] often [illegible] [illegible]

process [illegible] [illegible] [illegible]

as used [illegible] [illegible]

trunk or a [illegible] [illegible] [illegible]

Correlation [illegible] [illegible] [illegible]

device [illegible] [illegible] [illegible] [illegible] [illegible] a [illegible]

ldity [illegible]

[illegible] by [illegible] [illegible]

and Paper.[illegible] [illegible] [illegible]

disjunction [illegible] [illegible] [illegible]

[illegible]

[illegible]

[

buddy rankings and ratings and the results of certain selection devices, namely, an officer candidate (high level intelligence measure) test, several personality measures, and ratings by former superiors. In addition, it was determined by factor analysis that all ratings were loaded with a heavy leadership factor.

The results led the authors to conclude that:

(a) "Buddy ratings appear to be the purest measure of leadership."

(b) Buddy nominations (ranking) "are equally good measures of leadership."

(c) "Nominations by class appear to better be measures of the leadership factor than any other variable."

(d) Buddy nominations correlated more highly with all the selection devices, except intelligence, than did academic grades.

(e) Buddy nominations and ratings both measure accurately after a shorter period of observation than ratings by supervisors. That is, one's peers form accurate judgements of one more quickly than do one's superiors.

An Army study of buddy rankings in Officer Candidate School at Fort Benning, Georgia yielded a split-halves reliability coefficient of +.90 for a class which had been in being for five weeks and +.72 for a class of three weeks.[16] Correlations between the buddy rankings and the platoon officers'

Buddy ranking and rating and the results of certain selection devices, namely, an officer candidate (high level intelligence measure) test, several personality measures, and rating by former superiors. In addition, it was determined by factor analysis that all rating were loaded with a heavy leadership factor.

The results led the authors to conclude that:

(a) "Buddy rating appear to be the purest measure of leadership."

(b) Buddy nominations (ranking) "are equally good measures of leadership."

(c) "Nominations by class appear to be better be measure of the leadership factor than any other variable."

(d) Peer participation correlated very [illegible] with all the selection devices, except intelligence, thus did measure greater.

(e) Buddy nominations and rating both measure somewhat differently than peer participation or nomination than rating by superiors. That is, peers have more accurate judgments of an [illegible] than do their superiors.

In a study about [illegible] Buddy ranking in officer candidate schools at Fort Benning, Georgia, [illegible] a multivariate reliability coefficient [illegible] for a class of [illegible] in the [illegible] year. It has a class of three factors, [illegible] [illegible] methods [illegible] and [illegible] officer

rankings were as follows:

(a) For the five weeks class +.70.

(b) For the three weeks class +.65.

These figures represent rather close agreement and perhaps may be explained by the fact that the process required that the raters rank only the five best and the five poorest man in each class. This eliminated the middle group where greatest disagreement as to the correct order is likely to occur.

Another defender of ratings by peers is Bittner.[17] He argued that ratings by peers are likely to be more accurate because:

(a) Men live and work much closer to their peers than to their superiors and thus have a better opportunity to observe each other's performance.

(b) Men tend to show only their best sides to their superiors whereas one's peers have an opportunity to observe typical performance.

(c) The investigator can employ the principle of obtaining multiple judgements on the ratee more effectively if he uses buddy ratings.

Statements (b) and (c) apply equally well to ratings by one's juniors though there may be some question as to whether one's subordinates are able to judge the value of typical performance as well as one's peers can judge it.

<u>Ratings by peers instead of superiors.</u> The superiority of buddy ratings demonstrated by Therry and Fryer and

rankings were as follows:

(a) For the five weeks class +.70.

(b) For the three week class +.60.

These figures represent rather close agreement and perhaps may be explained by the fact that the process required that the raters rank only the five best and the five poorest men in each class. This eliminates the middle group where greatest disagreement of the common order is likely to occur.

Another reference of interest is found in Ulrich's letter.[17] He argued that raters by peers are likely to be more reliable because:

(a) men live and work much closer to their peers than to their superiors and thus have a better opportunity to observe each other's performance.

(b) men tend to know only their peers and to their superiors whereas one's peers have an opportunity to observe typical performance.

(c) The investigator can employ the principle of obtaining multiple judgments on the ratee more effectively if he uses only raters.

Statements (b) and (c) apply equally well to ratings of one's juniors though there is the same question as to whether one's subordinates are able to judge the value of typical performance as well as one's peers know it.

Ratings by peers between 1 supervisors. The experiment of Kelly and Fiske and Kelly's study within Westinghouse of Harry and Fiske and

hypothesized by Bittner raises the question as to why we should not dispense with ratings by superiors and use instead ratings by peers exclusively. Bittner answered this question in favor of retaining the custom of ratings by superiors when he wrote that, as was mentioned in the preceding chapter, 78% of the 1800 Army officers polled on the subject preferred not to be rated by their equals.[18]

The reasons given by these officers were not reported but one may speculate as to their nature. It seems probable that, if an organization obtained buddy ratings as a matter of routine, the consequent feeling of being continuously watched and judged by friends and fellow workers would generate undesirable tensions among the personnel of the organization. These tensions might easily prove to be so detrimental to morale and teamwork that the ensuing loss of efficiency would outweigh any advantage gained from the superiority of the buddy rating as a rating method.

<u>Future possibilities.</u> The complexity of the activities and duties of the executive has been suggested and the necessity for measuring executive performance largely by judgement has been stated and deplored. What then is the solution? Is there any likelihood that anyt ing can be done to aid significantly in the task of judgement? Is there any possibility of a superior substitute for judgement?

One possibility of a strictly Naval application is suggested by information in a letter from Dr. Everett G. Brun-

hypothesized by Lehner raises questions as to why we should not disagree with ratings by superiors and use instead ratings by peers exclusively. Lehner answered this question in favor of retaining the custom of rating by superiors when he wrote that, as was mentioned in the preceeding chapter, "the of the IPOC Army officers polled on the project preferred not to be rated by their equals."[16]

The reasons given by these officers were not reported but one may speculate on their nature. It seems probable that, if an impersonal matter [illegible] or a matter of rankling, [illegible] [illegible] [illegible]

dage of the Bureau of Naval Personnel recently received by the writer.[19] It is stated in this communication that data taken from the readily quantifiable items of all the fitness reports of one Naval Academy class of about thirty years experience as commissioned Naval officers show, that the Navy Fitness Report has "fairly good validity" in terms of predicting success before the Navy selection boards.[20] In addition, the data show that reports made on officers when they are holding the relatively junior ranks of Lieutenant (junior grade) or Lieutenant, are more accurately predictive of the subsequent attainment of flag rank[21] than are the fitness reports received while the officers are in the higher ranks. Information of this kind could not be used to validate any completely new technique or method. However, detailed analysis of it might yield information which would be of assistance in an evaluation of past practices made with a view toward synthesis of the best portions of earlier fitness report forms into a new device.

Other possibilities lie in the direction of determining and considering quantitatively more of the factors which are indicative of satisfactory executive performance. One of the important indicators of the quality of executive performance is the morale and attitudes of the rank and file of the organization. It is conceivable that if information as to the morale and attitudes of the subordinates of the officers composing the criterion group were considered in addition to the buddy ratings of the criterion group, a more valid criterion might be

dage of the Bureau of Naval Personnel research received by the
writer.[59] It is stated in this communication that data taken
from the Manual quantitative scores of all the fitness reports
of one Naval Academy class of about twenty years experience as
commissioned Naval officers exist, the Navy fitness report
has "total, and validity" in terms of predicting success be-
fore the Navy selection board.[60] In addition, the data show
that reports have no utility even as a method of telling the mili-
tarily junior ranks of "promotion" (junior grade) or disenrollment,
are more accurately predictive of the "success" attainment of
flag rank[61] than are the [illegible] "successful" officers
officers are in the lower ranks. Information of this kind
could not be used in validating the [illegible] method
method. However, detailing exactly [illegible] this infor-
mation would be a research [illegible] implication of test
procedure use with a [illegible] (Dunnette) [illegible]
[illegible] of earlier fitness [illegible] a new [illegible].

Other possibilities lie in the direction of [illegible]-
ing and correlating quantitative [illegible] of the feature of
the validity of [illegible] [illegible].
The [illegible] indicators [illegible]
is the [illegible] relationship [illegible]
[illegible]. It is [illegible] [illegible]
[illegible] [illegible]
[illegible] [illegible]
[illegible] [illegible]

achieved.

An ambitious effort toward determining the factors which are indicative of the quality of executive performance is the ten year project of leadership studies currently in progress on the campus of The Ohio State University.[22] The Personnel Research Board of the University is conducting the project and has given a statement of its purpose which is quoted in part below:[23]

> The purpose of such studies is to develop facts and methods which may be useful in the selection and training of persons for positions of leadership...

This group has made studies of the personnel and organizations of Naval shore establishments and a number of commercial firms and it plans to continue with detailed studies of the personnel and organization of a considerable number of small Naval vessels and aircraft squadrons. Small vessels have been chosen as fields for study because a greater number of comparable units will be available. Briefly stated the objectives of these investigations are to find out,(1) what leaders do,(2) what effect leaders have on their organizations, and,(3) what elements, other than the leaders' qualities have affected the organizations.

If the project is successful, it should do much toward identifying and isolating the qualities which make for good executive ability and in providing improved means for measuring those qualities. An additional or concomitant product of

the program under these circumstances would be accurate objective criteria of executive performance.

the program under these circumstances would be accurate objective criteria of executive performance.

PART II

Chapter 4

SERVICE RATING IN THE ARMED SERVICES

Some Early Navy Forms

Information is not available as to exactly when the formal written service rating was first put into use in the Navy. However, the 1891 Report on the Fitness of Officers shown in Appendix X is an example of one of the earliest Navy forms. It constitutes cute evidence that Naval authorities have long felt the need for a means of systematizing and recording the senior's judgements of his subordinates. It is interesting, and perhaps a little discouraging, to note the fundamental similarity between this pioneer form and the fitness report now in use in the Navy. It indicates that progress in this field has been slow indeed.

Subsequent developments. The early Report of Fitness of Officers proved unsatisfactory and changes were made from time to time. By 1912 the fitness report form had grown to a length of four double-spaced typewritten pages and required the recording of considerably more in the way of detailed and diversified information. The 1912 form was an evaluational type, although the evaluational terms prescribed for use had been cut down from the "Excellent", "Good", "Tolerable", and "Not good" of the 1891 form to a list of words, each of which was to be further qualified by numerals. For example, one could not simply rate an officer "Excellent" on any given trait,

PART II

Chapter 4

[illegible]

The Early Navy Years

Information is not available as to exactly when the formal written service rules, [illegible] in the Navy. However, the [illegible] report as was filmed by [illegible] shows in Appendix [illegible] the [illegible] Navy Journal. It contains [illegible] [illegible] tion in a long Tactics and [illegible] [illegible] recording the analysis [illegible] [illegible]. It is interesting [illegible] the fundamental [illegible] [illegible] tem report was in use [illegible] [illegible] years in the Navy.

Tactical Developments. [illegible] Leader in Training of Officers [illegible] [illegible] time on [illegible] [illegible] to a length [illegible] [illegible] the [illegible] and diversified [illegible]. The [illegible] type, [illegible] [illegible] bee [illegible] "Pol[illegible] [illegible] was [illegible] [illegible] co[illegible]

but must qualify the "Excellent" by adding the appropriate numeral between "3.5" and "4.0". The qualifying numerals to be used with "Very Good" were "3.0 to 3.5" with lesser numerals down to and including "0" assigned for use with the less favorable evaluative terms.

In a letter to the writer, Dr. Everett O. Brundage of the Bureau of Naval Personnel wrote:[1]

> ...These records show that not many major changes occurred after 1912. At this time numerical ratings on a scale of 0.0 to 4.0 were already used. A general "aptitude for the service" rating was also in use, and so were the questions on attitude toward the officer's being "under your command" and toward his being promoted. Later changes in these questions were toward increased quantification. The number of specific ratings [traits] varied after 1917 from 9 in 1919 to 18 in 1921-1924, decreasing to 14 in 1928. Note that in 1932 there were different [rating] forms for sea and shore duty. In general, the series of format changes in the reports indicate an attempt to force more accurate and objective ratings.

The Present Navy Form

OFFICER'S FITNESS REPORT NAVPERS-310A (REV. 6-45) is the formidable title of the rating device in current use in the Navy. As shown in Appendix D it is principally a graphic scale with normal distribution categories superimposed upon it. It cannot be classed as a forced distribution scale, however, as the instructions in re the distribution are merely hortatory not mandatory. Space for the usual biographical data is provided at the top of the sheet together with space

but must qualify the "excellent" by adding the appropriate numeral between "3.5" and "4.0". The qualifying numerals to be used with "very good" were "3.0 to 3.5" with lower numerals lower to not qualifying "0" rating for use with the less favorable evaluative terms.

In a letter to [illegible], [illegible] of the Bureau of Naval Personnel wrote:

> ...[illegible] [illegible] 1.0 [illegible] [illegible] [illegible]

for the ratee to indicate his preference for next duty, and further down the page, for the reporting senior to make his recommendations on that subject.

Also located in this top section is a requirement for the ratee to indicate his knowledge, or lack of it, as to whether he is qualified for sea duty. This refers to physical qualifications. This estimate by a layman, and not a disinterested layman at that, is of little value, in the opinion of the writer. Except when the officer has been taken ill suddenly and has not had an opportunity to seek the help of a Naval medical officer, information more reliable than his opinion is available. Rating scale authorities generally agree, it will be recalled, that qualities determinable by more objective means should not be listed on rating scales.

The next section provides ample space for reporting any particualr qualifications which the reporting senior believes should be given emphasis. Note that here again a layman's opinion on physical fitness is required.

Looking closely at the rating scale portion of the form we see that it is abbreviated somewhat and is not completely of the descriptive type. Moreover, it has no provisions for minimizing halo effect. Earlier discussion in this thesis has suggested the value of longer scales and has shown the superiority of the descriptive scale and the need for minimizing halo.

Below the rating scale is a section designed to elicit

for the rates to [illegible] preferences for [illegible], and further down the [illegible] the proposition [illegible] to make his recommendations on [illegible] subject.

Also there is in this [illegible] a requirement for the rates to increase in [illegible] of it, or to whether he is qualified for [illegible]. This refers to physical qualifications. This applies to [illegible] and such a distinctness [illegible] at [illegible] of little value, [illegible] opinion of the writers. [illegible] ill health and [illegible] opportunity to [illegible] of a Naval medical officer, [illegible] his opinion is available. [illegible] agree, it will be [illegible], but [illegible] more objectively [illegible] or rating [illegible].

The next section [illegible] for [illegible] any particular qualifications [illegible] believe [illegible] many [illegible] of [illegible].

[illegible] closely [illegible] we see that [illegible] of [illegible] minimum [illegible] suggestion [illegible] level of [illegible] [illegible] [illegible].

an indication of the reporting senior's attitude toward having the ratee under his command. The next section below this is designed to determine his feelings toward promoting the ratee. There is also some attempt made toward evaluating the entire fitness report by securing a rough estimate of the number of officers considered by the rater in setting his standards.

Proceeding with the examination we find a provision for citation of commendatory or adverse material, space for a general overall evaluation, and a space in which the rater must indicate whether or not the ratee has seen the report.

Physical makeup. The complete form is made up in a packet of four sheets with carbon paper between the second and third and between the third and fourth sheets. The first page carries on its face the instructions to the reporting senior and on its back a replica of the complete rating form to be used as a rough work sheet in making the rating. The second sheet also has a complete rating form on it while the third has only sections 1 through 6 and the fourth page shows sections 1 through 5. All sheets require the signatures of both the rater and the ratee.

Use of form. When the rater is satisfied with the rating he has made on the worksheet, it is transcribed by typewriter to the second, third, and fourth sheets and signatures are affixed. (It is customary, though not essential, for the ratee to sign the forms before the rating is made since his signature certifies only to the correctness of the information

an indication of the reporting officer's attitude toward giving the ratee under his command. The next section [illegible] scale is designed to determine his feeling [illegible] the issue. There is also some attempt made toward evaluating the outline fitness report by asking a number of questions [illegible] officers concerned by the ratee in question. [illegible]

Proceeding with the examination of the [illegible] a provision for obtaining of comments by [illegible] officers for [illegible] overall evaluations and a space [illegible] and indicate whether or not [illegible].

Physical appearance. The [illegible] pocket [illegible] and chest and between the [illegible] page carries on the face [illegible] senior and on its back [illegible] to be done as a [illegible] second sheet also has a [illegible] third [illegible] section [illegible] both [illegible] and [illegible].

Use of [illegible]. [illegible] in [illegible] walker [illegible] are [illegible] [illegible]

he has contributed to the form.) The next step is the mailing of the second and third sheets to the Bureau of Naval Personnel for such processing, filing, and subsequent evaluation or consultation as may be required. The fourth page becomes the property of the ratee when his report is mailed to headquarters by the rater.

Some analyses of Navy ratings. An extensive study was made in 1945 by the Training Standards and Curriculum Division of the Bureau of Naval Personnel:[2]

> ...to evaluate the Report on the Fitness of Officers as a possible criterion measure for use in validating officer selection and classification requirements.

The fitness report files of 534 junior officers who had entered commissioned service between December 1940 and December 1942 were examined. From each file information was selected from two fitness reports, one from each of two different reporting seniors. Correlations were obtained between the various items of this information. The more significant for our purposes are given in the following table.

Coefficients of reliability between marks by the two seniors:	Service throughout the reporting periods.
From +.22 to +.48	Sea duty both periods.
From +.0004 to +.18	Sea duty one period, shore duty the other.

From the above, the unreliability of the ratings is obvious. These figures were obtained from ratings on Navy forms

the low correlations on the forms. The next step in the qualifying of the second set of items to the season of level personnel for such procedures utility, so subsequent evaluation as correlations of low or negative. Therefore, a low degree the presence of the same man for marks is willing to broadcasters of low ability.

Item analysis of test criteria. In a similar study was made in 1941 by the Training Department of Navigation of the Service of Transport Canada.

...an analysis of the results of the Figures of Difficulty for a set of 100 items was made in validation of the Aptitude Test. The Figures of Difficulty were calculated for each item and plotted against the Figures of Difficulty for the same items as calculated for two Figures of Difficulty were calculated for each item and plotted against the Figures of Difficulty for the same items as calculated for two groups of men, one group being composed of men who had failed the Aptitude Test.

Coefficient of reliability between two groups	Positive correlation between two groups
[illegible]	[illegible]
[illegible]	[illegible]

now superseded but, due to the fact that the currently used Navy scale is less of a descriptive scale than former ones were, it is doubtful if the current scale would yield significantly better results.

Cagle reported results of a study of the distribution of marks obtained on the current Navy fitness report form. His figures, reproduced below, reveal a marked divergence from the normal distribution.[3]

Percentage Group	Correct (Normal) distribution	Actual distribution	Actual Percentage
Top 10%	119	494	41.6
Next 20%	238	578	48.7
Middle 40%	474	102	8.6
Lower 20%	238	12	1.0
Bottom 10%	119	2	0.0
Totals	1188	1188	99.9

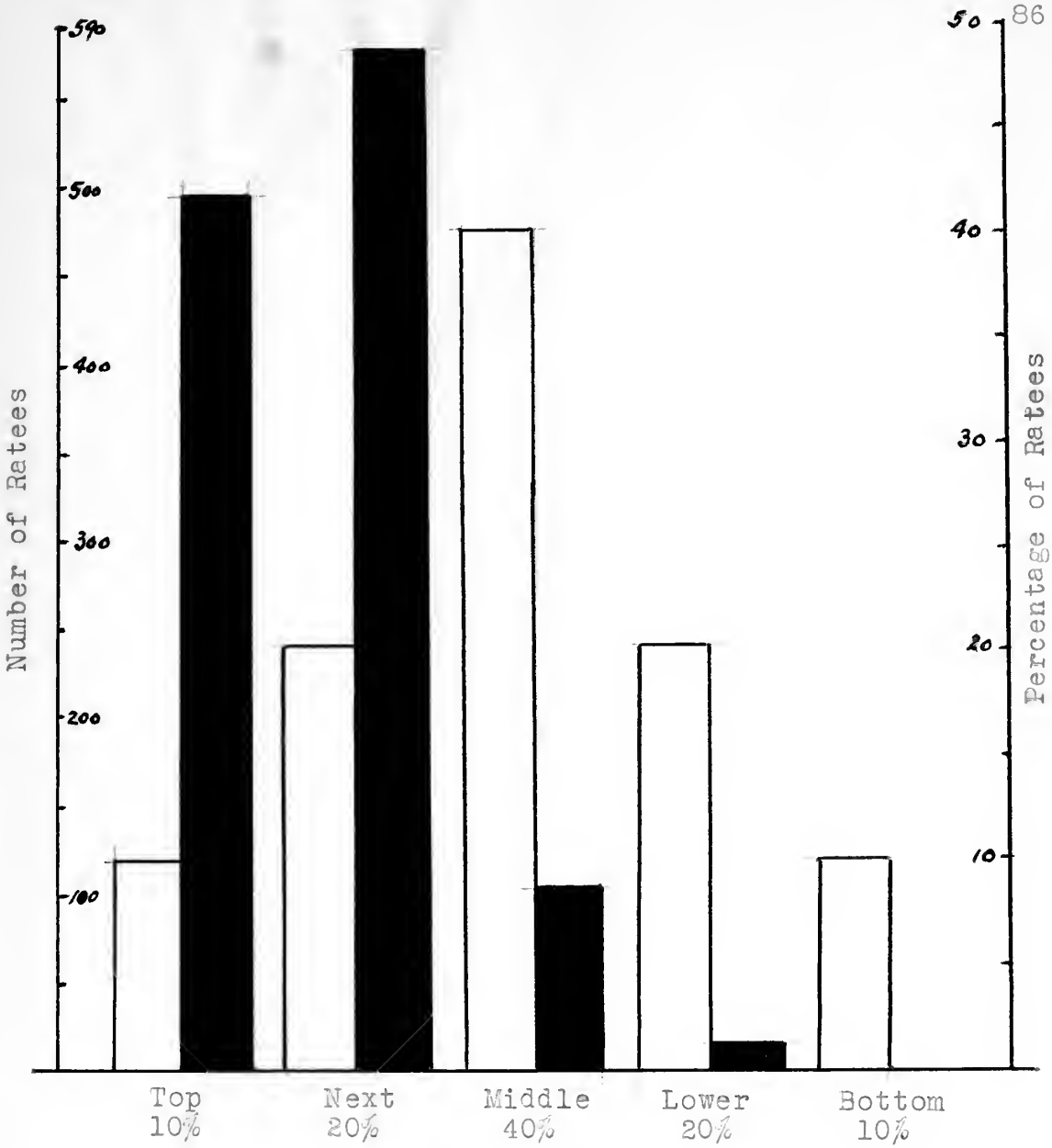
A similar study of the ratings made with the current Navy form on Forty-seven Ground School Instructors at the Naval Training Command at Pensacola, Florida was reported by Richardson, Bellows, Henry and Company, Inc., in 1948.[4] Their figures are as follows.

now superseded by, in the last line, the currently used Navy scale in Area 2. Apparently some other work, it is doubtful if the current scale will yield significantly better results.

Cagle reported results of a study of the distribution of marks obtained on the current Navy Lifter report form. His figures, reproduced below, reveal a marked divergence from the normal distribution.[2]

Percentage Group	Normal (Normal) Distribution	Actual Distribution	Actual Percentage
Top 10%	110	404	41.6
Next 20%	220	378	44.7
Middle 40%	440	132	9.6
Lower 20%	220	10	1.0
Bottom 10%	110	2	0.2
Totals	1100	1100	99.1

A study of the method used by the Navy for determining Navy Jobs of Larger Value found great shortcomings at the Naval Training Center, San Diego, Calif. as pointed out by Mr. Gibson, Jones, and Mr. Cochran, Navy, in the.[3] It is felt that the same is true elsewhere.



Unshaded Bars - Prescribed Distribution
Solid Bars - Actual Distribution

Figure 4

GRAPHIC PRESENTATION OF DISTRIBUTION DATA
OBTAINED FROM STUDY OF NAVY OFFICER RATINGS

Percentage Group.	Actual Percentage Distribution Found.
Top 10%	56%
Next 20%	44%
Middle 40%	2%
Next 20%	2%
Bottom 10%	0%
	104%

(Note: The error of 4% in the total results from an error introduced in rounding off the small numbers involved when computing percentages.)

These figures also show a marked divergence from the normal distribution.

There has been little done on validity studies of the Navy fitness report system other than as mentioned in the chapter on Criteria, page 77.

Evaluation. In favor of the Navy's form it may be said that:

(a) The instructions to the raters are brief and readily available.

(b) The form is designed to reduce the labor of preparation to a minimum.

(c) The form has provisions for recording all information pertinent to the permanent service record of the rates except that which is determinable by more objective means (however, see items (c) and (d) below).

Percentage group.	:	Actual percentage [illegible].
Top 10	:	[illegible]
Next 20	:	[illegible]
Middle 40%	:	[illegible]
Next 20%	:	[illegible]
Bottom 10	:	[illegible]
		100%

(Note: The [illegible] to 40 of the [illegible] results from [illegible] [illegible] but the [illegible] [illegible] percentages.

There [illegible] also a [illegible] distribution.

[illegible] Next [illegible] [illegible] [illegible]

(d) The form presents the strengths and weaknesses of the ratee in an analytic and easily understandable form.

The Navy's rating system leaves much to be desired, however. Chief among such considerations are:

(a) No provision is made for rater training, an omission for which the excellent instruction sheet cannot compensate fully.

(b) The rating relies on judgement.

(c) The design of the form does not permit rapid and accurate sorting, tabulation, and evaluation of the data by machine.

(d) The form attempts to elicit information on physical fitness which is more accurately determinable by other means.

(e) The form gives precious space to the recording of the ratee's preference for next duty, a bit of information also furnished to the Bureau by the ratee annually, or more often if preferences change, on the OFFICER'S DATA CARD NAVPERS-340 (rev. 7-46).

(f) The rating scale is abbreviated and not completely of the descriptive type (in this respect, the rating scales of earlier Navy forms as shown in Appendix C were superior to the presently used one) and it provides no safeguards against the halo effect or rater's bias.

(g) The form makes no provision for assisting the reporting senior in preparation for conduct of the post-rating

(d) The form presents the [illegible] of the rotor in an analytic and easily understandable form.

The navy's radio, [illegible] however. Could [illegible]

(a) No provision is [illegible] an explanation for which the [illegible] are fully.

(b) The [illegible]

(c) The design of [illegible] accurate setting, [illegible] evaluation of the use of machine.

(d) The [illegible] cal figures which [illegible] means.

(e) The [illegible] the method [illegible] also [illegible] other [illegible]

NAVAL-SHIP [illegible]

(f) [illegible] of the [illegible] earlier [illegible] to the present [illegible] in the basic [illegible]

(g) [illegible] purpose [illegible]

interview with the ratee (see page 66).

(h) The rating procedure does not provide for the rating being made by the superior who has the most complete knowledge of the ratee. Instead, this is the prerogative of the commanding officer who, in a large commend, may have only the most superficial knowledge of the ratee.

(i) The rating procedure and the form make insufficient provision for the difference in duties and qualities required of officers, depending on whether they are of the Line or Staff, whether they are specialists within those categories, or whether they are in school or engaged in the duties of their respective classifications.

(j) The procedure makes no provision for correcting or adjusting the markings to compensate for the tendency of some raters to mark high and others to mark low.

(k) The rating procedure makes no provision for either the averaging of the ratings made by several raters or the review at higher echelons of ratings made at lower echelons.

(l) There is no provision for rating the rater.

(m) And, finally, the Navy's rating form has low reliability and produces an excessively skewed distribution of scores.

The Army Rating Form

As has been mentioned, the Army uses a form known as the "forced choice" ratin form.[5] Samples are shown in Appendix E. The most conspicuous features of this device are,

interview with the ratee (see page 68).

(h) The rating procedure does not provide for the rating being made by the supervisor who has the most complete knowledge of the ratee. Instead, this is the responsibility of the commanding officer who, in a large command, may have only the most superficial knowledge of the ratee.

(i) The rating procedure [illegible] provision for the difference in duties and responsibilities [illegible] of officers, depending on whether they are on the line or staff, whether they are specialists [illegible] other categories, or whether they are in general [illegible] the duties of their respective officerships.

(j) The procedure makes no provision for promotion or adjustment [illegible] the rating [illegible] the [illegible] of more raters to make it [illegible] other [illegible].

(k) The rating procedure makes no provision for giving the average of the rating [illegible] numbers of the reviewer at [illegible] or lower [illegible].

(l) There is no provision [illegible] later.

(m) And, finally, the [illegible] rating [illegible] reliability [illegible] is necessary, [illegible] of scores.

The Navy [illegible]

[illegible],

the "[illegible] rating [illegible]

[illegible].

(a) the forced choice items and, (b) the provision for performing the marking directly on a machine scorable sheet.

Looking at the sheet, we find it entitled EFFICIENCY REPORT with the additional identification symbols at the bottom of the page of WD - AGO FORM 67-1, 1 Jul 47. The upper half of the first sheet is occupied by the biographical section, followed about midpage by Section B, a space for a general overall evaluational statement by both the rater and the rating reviewer or indorser. Next, in Section C, is a list of nine important situations with space for the rating officer and the indorsing officer to indicate their attitudes toward having the ratee serve in them. Following this are spaces for recommendations by the rater and the indorser for future duty assignments and future training. Lastly there is a device intended to determine what experience with the ratee the rater used as a basis for his judgements.

The reverse side of the first sheet is blank. It is reserved for the use of the Career Sections of the Arms and Services of the Department of the Army.

The biographical data are repeated on the second sheet followed by the first section of forced choice items, Section IV. These items are intended to measure the rater's judgement of the ratee's job proficiency. Section V of this sheet, also entitled Job Proficiency, is designed to secure judgements from rater and indorser as to which of a list of seven specific positions the ratee is, or would be, most proficient

(a) the forced smoke item, and, (b) the provision for performing the washing directly on a washing accessible sheet.

Looking at the sheet, we find it entitled PRELIMINARY REPORT with the conditional identification symbols at the bottom of the page of "R - ARC POW QT-1, I ARC 47". The upper half of the first sheet is separated by a horizontal section, followed by a subsection B, a space for a general overall evaluation and recommendations by both the rater and the rating officer or in person. Then, in section C, is a list of nine important characteristics with space for the rating officer and the immediate officer to indicate the assistance toward having the rater move to them. Following this is a space for recommendations of the rater and the immediate for future duty assignments as Command Pilots. Lastly, there is a section intended for interview and discussions with the rater to be used as a basis for this information.

The reverse side of the first sheet is blank. It is reserved for the use of the Career Sections of the Army and Navigator of the Department of the Air Force.

The objectives have been revealed by the rating sheet followed on the first section of forced choice items, section IV. [illegible]

[illegible]

in and which is the second most suitable job from the point of view of his proficiency in it.

On the reverse side of the second sheet there are two sections designed to determine the personal qualifications of the ratee. The first of these, Section VI, consists of forced choice items. The other, Section VII, lists six personal qualification items with spaces for the rater and indorser to indicate the degree to which the ratee exhibits each.

Finally, there is an attempt to determine where the ratee stands, in the opinion of the rater, relative to the other officers of his rank who have been rated at this time by the rater.

Physical makeup and use. The Army rating form is a leaflet of two pages with dimensions of 9 inches by 11 inches and arranged to open from the right edge like a book. It is not equipped with carbon paper but its construction makes it easy to insert a piece when typing in the biographical data, the only entries which require duplication. As mentioned earlier, the sheets are machine scorable. Thus, if the marking is performed with an "electrographic pencil" as prescribed, the reports may be scored by machine without further processing.

There is provision for the signatures of both the rater and the indorsing officer but not of the ratee. Moreover, there is no requirement that he be shown his rating. It is true that the meaning of the markings on the forced choice sections will be known only to the possessor of the scoring key in the

in and which is the second most suitable job from the point of view of his proficiency in it.

On the reverse side of the second sheet there are two sections designed to determine the personal qualifications of the raters. The first of these, section VI, consists of forced choice items. The other, section VII, lists six personal qualification items with spaces for the rater to indicate to indicate the degree to which the ratee exhibits each.

Finally, there is an element in determining where the rates stands, in the opinion of the rater, relative to the other officers of his rank who have been rated by the rater.

Physical rating and use. The rating form is a last-last of two pages with dimensions [illegible] of 11 inches and arranged to open from the right as [illegible] a book. It is not equipped with carbon paper but [illegible] makes it easy to insert a piece when typing in the biographical data, the only exception [illegible] duplication. [illegible] the ratees are [illegible] alphabetically. Thus, in the middle [illegible] formed with an "alphabetically [illegible]", the reports [illegible] without further processing.

There is [illegible] the [illegible] and the number and the [illegible] of the raters. [illegible] there is no [illegible] [illegible]. It is [illegible] that [illegible] will [illegible] [illegible]

Adjutant General's Office, but certainly the ratee could profit by seeing the ratings made on Sections C, V, and VII as these are merely graphic rating scales.

The forced choice theory. Sisson rather thoroughly described and discussed forced choice items, the theory underlying them, and the methods employed in their construction.[6] Only a brief treatment will be attempted here.

Developers of the forced choice technique reasoned that if items could be offered to the respondent in pairs, so chosen that each appeared equally acceptable or unacceptable, and further, if these items could be so selected that choice of one would be an indication of the presence of one quality or opinion and choice of the other would indicate another quality or opinion, then they would have a device for measuring qualities or opinions which could not be controlled by the respondent. The fact that although both items might appear to the respondent to be irrelevant, he must nevertheless make a choice, gave the method its name.

To improve the rater's attitude toward the device, the Army rating designers catered to the desire of people to speak well of others if possible, a desire which leads to a tendency in rating known as "generosity error". They arranged their items in tetrads of two favorable and two unfavorable statements with the provision that the rater must choose only one statement which is most like the ratee and a second which is least like him.

Adjutant General's office, but unfortunately I have been unable to profit by seeing the writings of Col. [illegible], and VII as these are merely pseudo-[illegible] [illegible].

The Moral Choice Theory. Under rather burdensomely described and licensed moral choice items, the theory underlying them, and the methods employed in their construction, only a brief treatment will be attempted here.

Analysis of the moral choice technique indicated that if items could be derived to the respondent in terms so chosen that some apparent moral choice would be unacceptable, rationalization, if there were such, would be supplied when choice of one or other [illegible] of the quality or opinion was stated in the other would indicate another quality or opinion, which [illegible] would have a better low measuring qualities or opinions which could not be contradicted by the respondent. The fact that they disclosed [illegible] appear to be reasonable to the respondent, he more [illegible] have made a choice, gave the method its name.

To improve the research military [illegible], the Army [illegible] designed a canvas to be [illegible] to [illegible] wall or screen if possible, a [illegible] [illegible] to a [illegible] in public, such as "[illegible]". [illegible] items in terms of [illegible] [illegible] [illegible]

The data as to the apparent favorableness or unfavorableness of the items and their discriminatory value are determined by empirical methods.

Some analyses of Army ratings. The extensive testing which was done in connection with the design of the Army rating form required nearly two years to complete and yielded results which led the investigators to conclude that the new form was the "most valid"[7] of those tested and further, it was "clearly superior"[8] to the older form it was designed to replace. It is unlikely that any organization, private or military, will attempt so extensive a project in the forseeable future. Consequently, the Army's rating system probably will retain the distinction of "the most thoroughly tested" for some time to come.

The criterion group was obtained by the sociometric method discussed in the preceding chapter and consisted of nearly 50,000 officers. These officers were rated on several different devices including the Army form then prescribed for official use, WD AGO FORM 67, and two versions of the new forced choice design. FORM 67 is described rather completely by Sisson but in general it was merely an evaluative graphic scale.[9] Correlations of the results prompted the conclusions mentioned above but none of the correlations were as high as those reported by Probst and by Knauft for their respective check list rating scales, (see preceding chapter). It is only fair to add, however, that the validities obtained by the Army

The data as to the opposing favorableness or unfavorableness of the items and their distributions, which are determined by empirical methods.

Some analyses of Army ratings. The extensive testing which was done in connection with the design of the Army rating form required nearly two years to complete and yielded results which lay the correlations to provide that the new form was the "most valid" of the - tested and further, it was "clearly superior" [illegible] It was designed to replace. It is unlikely [illegible] superiority, however or military, will attempt to [illegible] in the foreseeable future. Consequently, [illegible] agency probably will retain the discretion [illegible] "economically based" for some time to come.

The selection group [illegible] mental learned in the [illegible] of nearly 50,000 officers. [illegible] based on several different devices [illegible] official use [illegible] Forces [illegible] by [illegible] scale." [illegible]

investigators were based on enormously greater numbers of cases and that, whereas the Army experimenters used buddy ratings as the criterion, both Probst and Knauft used supervisors' judgements. It seems likely that supervisors' judgements would correlate more closely with ratings made by supervisors than with buddy ratings.

More recent studies of ratings obtained with the new Army form have yielded the distribution curves presented in Figures 5, 6, 7 on pages 96, 97, and 98 respectively.[10] Although as shown by Sisson,[11] the forced choice rating gives a slightly better approximation of the normal distribution than the FORM 67 which it replaced, these more recent studies reveal that it falls far short of attaining the ideal, normal distribution.

Evaluation. The currently used Army form has several points of superiority. Principal among these are:

(a) By "disguising" the nature of the marks being given, it minimizes the ability of a biased rater to make an unduly biased report.

(b) The form is designed to minimize the labor required to prepare it, to mark it, and to score it.

(c) It provides for review and indorsement by higher authority.

(d) The procedure provides that the rater's immediate superior will perform the rating.

(e) No attempt is made to measure qualities which may

[illegible]

[illegible]

be better determined by more objective means.

(f) The form provides a means to indicate the basis on which the rating was made, thus affording information to aid in evaluating the rating.

(g) The validity is superior to that of the evaluative rating scale.

(h) The distribution of scores produced by the forced choice device approximates the normal distribution more closely than does the distribution from the evaluative graphic scale. In particular, the forced choice form causes more ratings to fall into the "average" and "below average" zones and thereby discriminates better between officers in the lower end of the distribution.

The Army's rating system has a number of shortcomings too. The more important of these are:

(a) There is incomplete provision for an indication of the strong points and weaknesses of the ratee. Certain portions, Sections C, V, and VII, of the rating device do show some information of this kind but the two sections of forced choice items conceal their true indications from the ratee and rater.

(b) There is no provision for discussing the rating with the ratee or for certifying to him that the rating has been duly made and submitted to headquarters.

(c) There is no provision for rating the raters.

(d) The rating form and procedure make insufficient

be better determined by more objective means.

(f) The form provided seems to indicate the basis on which the rating was made, thus affording information to aid in evaluating the rating.

(g) The validity is superior to that of the evaluative rating scale.

(h) The distribution of scores resulted in the forced choice device being a proximate normal distribution more closely than does the distribution of the evaluative graphic scale. In particular, the forced choice device gives more ratings to fall into the "average" and "below average" scores and thereby distributes rater errors differently in the lower end of the distribution.

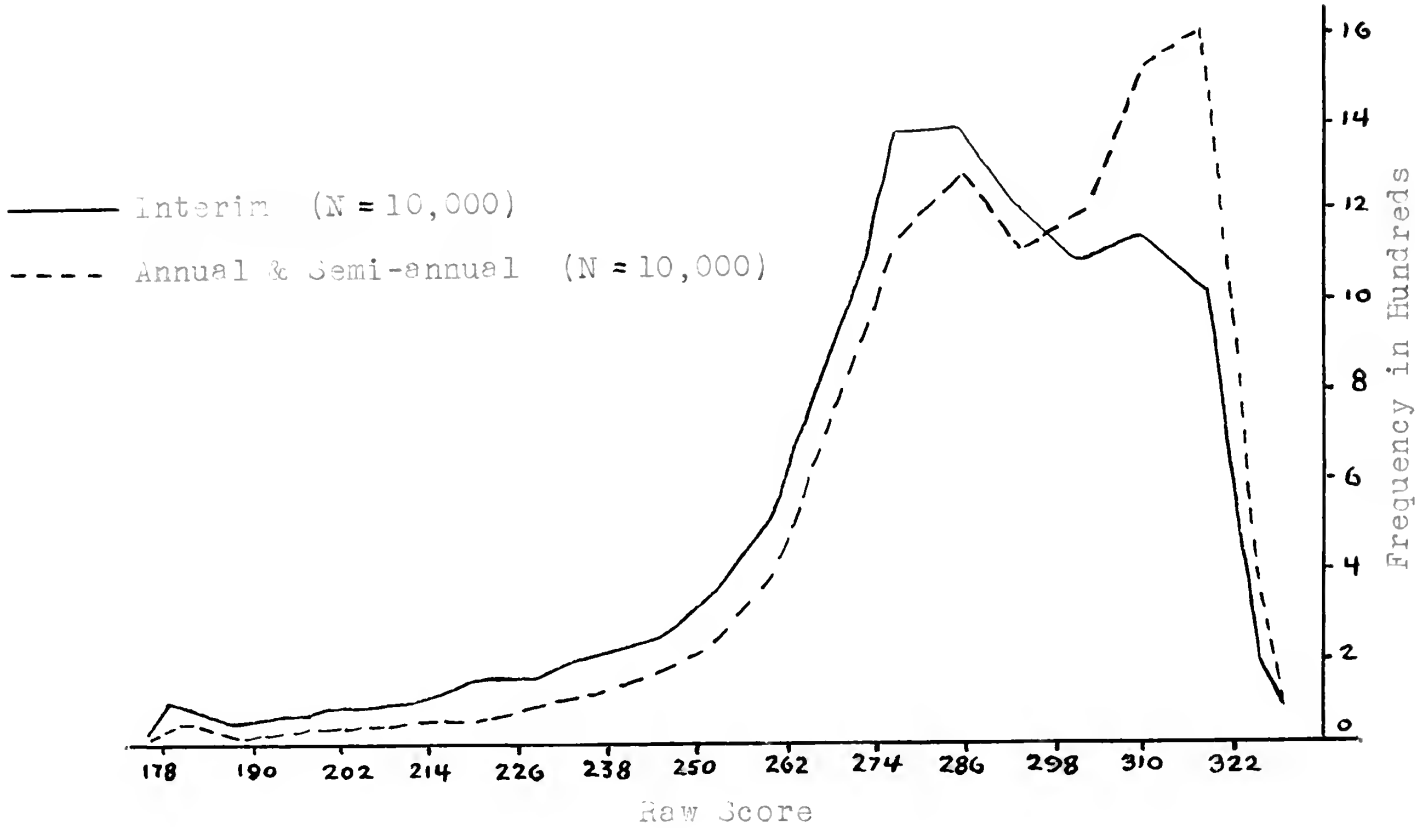
The simple rating [illegible] of subjectivity, too. The main importance of this type.

(a) There is considerable variation in the interpretation of the rating scale [illegible] too. Certain particulars, outline of, V, and VII, [illegible] do more [illegible] some information [illegible] forced choice device [illegible] the [illegible] and raters.

(b) There is no [illegible] [illegible] with [illegible] been [illegible] available to [illegible].

(c) There is no [illegible] [illegible].

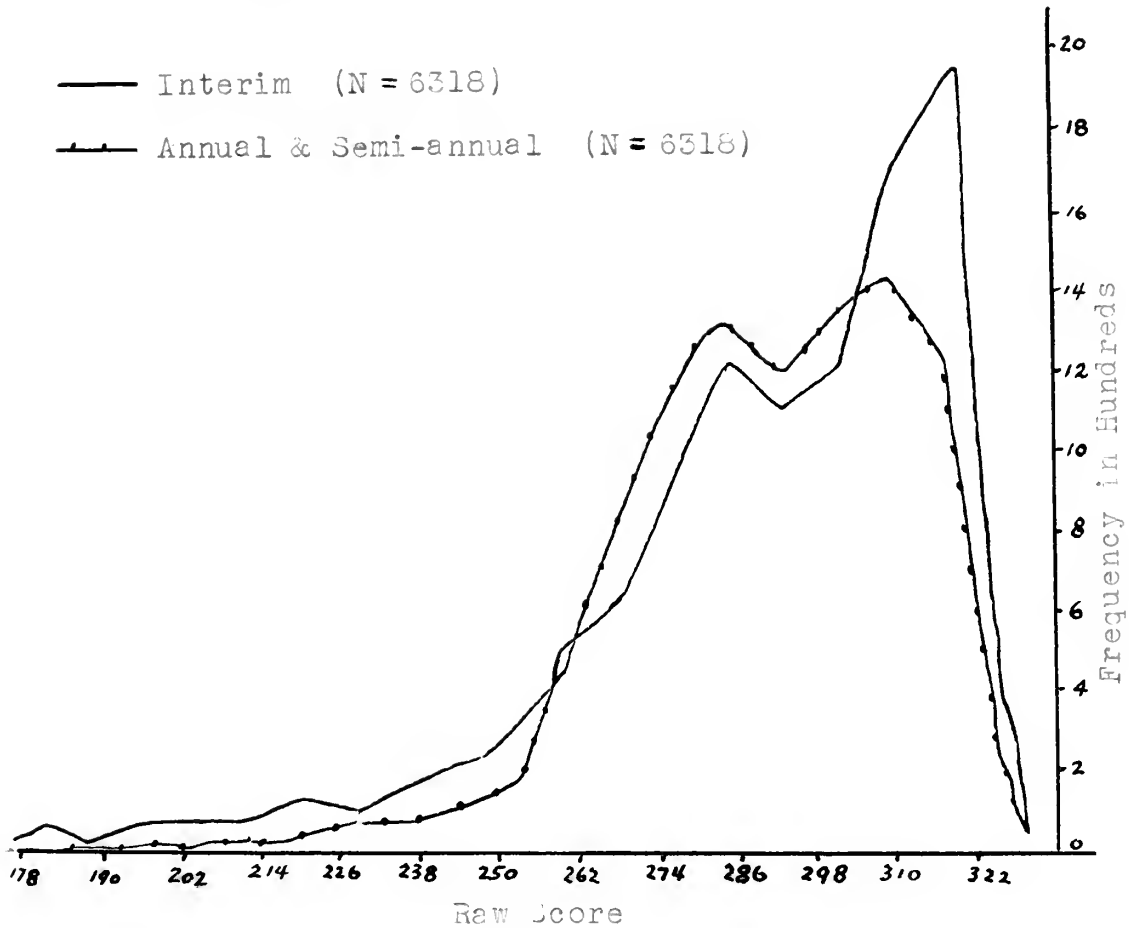
(d) The rating [illegible] procedure [illegible] in [illegible]



From AD AGO PRS REPORT 722, The 1948 Standardization of Officer Efficiency Report AD AGO FORM 67 - 1 7 June, 1949, p. 18

Figure 5

DISTRIBUTION OF OVERALL RAW SCORES ON ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1, ALL GRADES

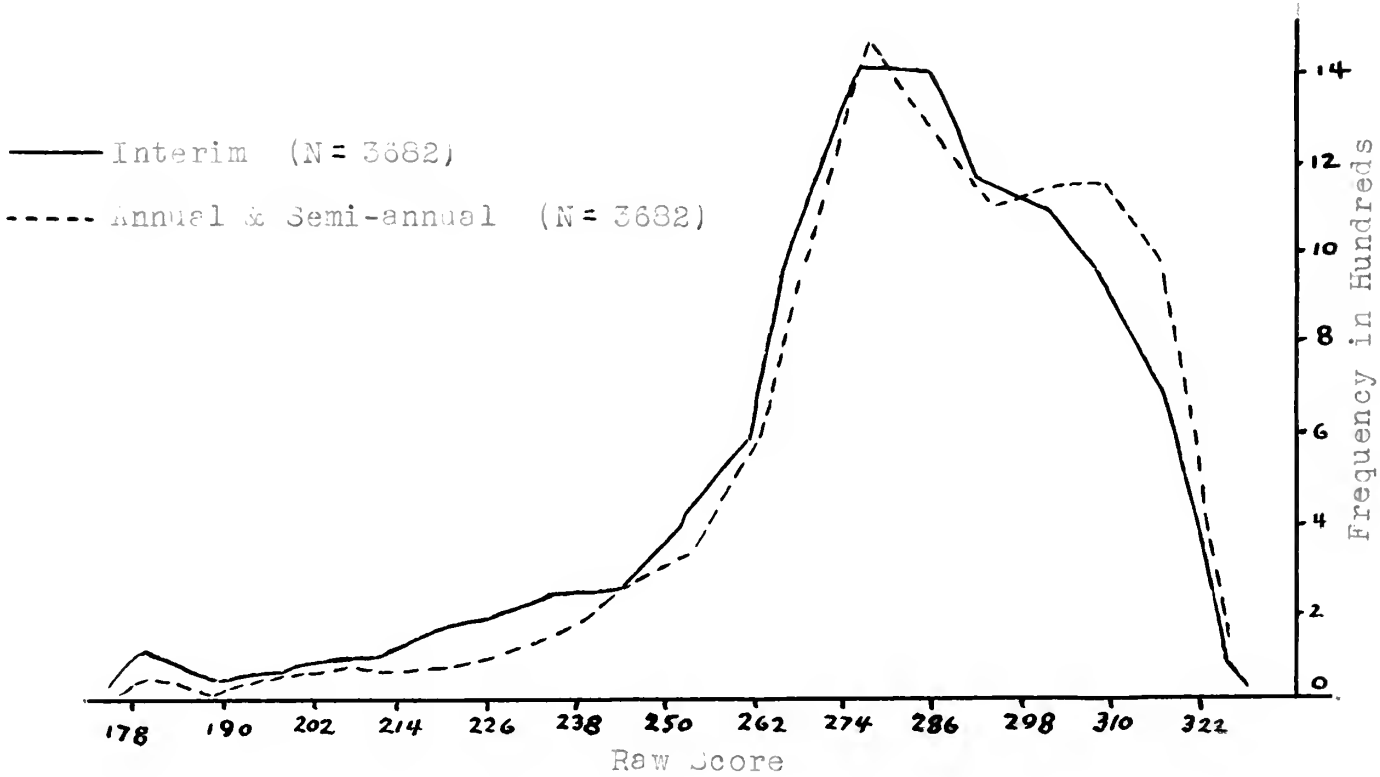


From WD AGO PRS REPORT 799, The 1948 Standardization of Officer Efficiency Report WD AGO FORM 67 - 1 7 June, 1949, p. 21

Figure 6

DISTRIBUTION OF OVERALL RAW SCORES ON ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1

Field Grade



From WD AGO PRS REPORT 799, The 1948 Standardization of Officer Efficiency Report WD AGO FORM 67 - 1
7 June, 1949, p. 22

Figure 7

DISTRIBUTION OF OVERALL RAW SCORES ON
ARMY OFFICER EFFICIENCY REPORT, FORM 67 - 1

Company Grade

provision for the difference in duties and qualities required of officers depending on whether they are of the Line or Staff, whether they are specialists within those categories, or whether they are in school or engaged in the duties of their respective qualifications. It is possible, however, that the empirically chosen forced choice items may automatically account for such differences.

The Air Force Report

The Air Force in 1948 adopted a new rating system which was the outcome of considerable research by the American Institute of Research of Pittsburgh.[12] It features a pamphlet or work booklet, in which the rater records the results of observations he makes of the ratee, and a single sheet rating form on which he summarizes the data recorded in the booklet. A sample page from the booklet is shown in Appendix F while the rating sheet is shown in Appendix G.

Description and use of booklet. To guide the rater in his observation of the ratee, the work booklet, entitled Check List of Critical Requirements for Officer Evaluation on Report of Officer Effectiveness, lists six areas in which proficiency, or the lack thereof, is significant. These six areas are:

I. Proficiency in Handling Administrative Details.

II. Proficiency in Supervising Personnel.

III. Proficiency in Planning and Directing Action.

IV. Acceptance of Organizational Responsibility.

provision for [illegible] of officers [illegible] whether they [illegible] or the [illegible] tive qualities [illegible] only [illegible] such differences.

[illegible]

[illegible] tude of [illegible] work [illegible] vantage [illegible] on [illegible] sample [illegible] rating [illegible]

[illegible]

[illegible]

V. Acceptance of Personal Responsibility.

VI. Proficiency in Duty Military Occupational Specialty.

Within each area of proficiency are listed specific observable behaviors, each divided into five degrees of competence with each degree indicated by a one digit numeral and a short descriptive statement. These specific observable behaviors are arranged in a column with the descriptive statements and numerals placed along a line opposite each as in a graphic rating scale. Thus if the rater, while considering the performance of the ratee in one of the significant areas, observes behavior which may be described by one of the phrases provided, he marks the digit over the statement. As this process continues throughout the period between ratings, a body of information about the proficiency of the ratee is built up based on his observed behavior. It is recorded in terms of the numerals with the figure "1" indicating minimum competence and "5" indicating maximum proficiency.

When the date for preparing the formal rating arrives, the marks in the booklet are transferred to spaces provided on the rating sheet which is a condensed graphic rating scale. The scoring is done by hand on this sheet, treating the marks with the prescribed weights.

<u>The rating sheet.</u> The rating sheet, which is entitled <u>Report of Officer Effectiveness,</u> provides space for data other than the summarized observations from the work booklet. There

V. [illegible]

VI. [illegible] in [illegible] development [illegible]-

ty.

[illegible] observable [illegible]

[illegible] force [illegible] short [illegible]

[illegible] behavior [illegible]

[illegible] monks [illegible]

[illegible] possible [illegible]

the [illegible]

observe [illegible]

provided, [illegible]

case [illegible]

of [illegible]

up [illegible]

of [illegible]

perhaps [illegible]

[illegible]

the [illegible]

or [illegible]

the [illegible]

air [illegible]

[illegible]

[illegible]

is, of course, a place for certain necessary biographical information. Next on the sheet is a rather large space (nearly one half of the page) for a general overall evaluational report. Next is a provision for the specific, formal delegation of the rating duty by the commanding officer to some other officer together with a statement of the reason for such delegation. This makes it possible officially to assign the reporting duty to the officer with the most adequate knowledge of the ratee. Following next is a place for the signature of the reporting officer (rater) and the comments and signature of the indorsing officer.

The reverse side of the sheet is given over to a restatement of the six significant areas of proficiency with space under each for entering the summarized numerical marks described in the preceding treatment of the booklet and its use. The spaces for the entries are so arranged that the scores may be calculated and summarized on the sheet and the total entered at the bottom. Lastly, there is a small space reserved for use in describing the ratee's "Military Occupational Specialty".

<u>Evaluation.</u> The strength of the system lies in its use of the work booklet to record actual observations of significant behaviors as they occur. It is, in effect, a comprehensive graphic rating scale of the descriptive type with the advantages inherent in such a device. Moreover, the booklet provides, to a considerable extent, for rating on the spot as

is, of course, a place for certain necessary bioregional information. Next on the sheet is a rather large space (nearly one half of the page) for a brief personal ornithological report. Next is a provision for the specific, formal delegation of the voting duty of the commanding officer to some other officer together with a statement of the reason for such delegation. This makes it possible officially to assign the reporting duty to the officer with the most adequate knowledge of the area. Following next is a place for the signature of the reporting officer (rater) and the commanding officer and signature of the inspecting officer.

The reverse side of the sheet is given over to a statement of the qualifications which a prospective member should possess, and for whatever the commanding officer considered in the preceding paragraph, the method and its use. [illegible] stated that the [illegible] approved as a guideline for [illegible] the sheet for the tools [illegible] research [illegible] subject "Military ornithological [illegible]".

<u>Conclusion.</u> In [illegible] of the system [illegible] in the use of the [illegible] ornithological [illegible] of military-[illegible] [illegible] [illegible] [

the behavior occurs, thereby eliminating the errors due to the frailties of human memory and securing a rating on the firm basis of actual observed behaviors.

The designers have reported a split halves reliability coefficient of +.89 using 534 ratings divided into equivalent halves. When corrected by the Spearman-Brown formula the reliability coefficient became +.94. This is the highest reliability coefficient for a rating device which the writer has found reported.[13]

Other advantages of this system include:

(a) Limitation of the rating to qualities which cannot be determined more accurately by more objective means.

(b) A design which permits quick and easy transfer of the information accumulated in the booklet to the rating sheet.

(c) The rating sheet is designed to present the markings in an analytic and easily understandable form.

(d) Provision is made for a general overall evaluational statement.

(e) Provision is also made for a statement of the ratee's specialty.

(f) There is provision for formal delegation of the rating responsibility.

(g) And finally, there is provision for the review and indorsement of the ratings.

Chief among the disadvantages of the system is the obviously unwieldy procedure of the rater's attempting to keep

the behavior occurs, thereby eliminating the errors due to the fragilities of human memory, and securing a reading on the firm basis of actual observed behaviors.

The ideal here is to develop a still higher reliability coefficient of 1.00 with the [illegible] "value" into an invalid behavior. This [illegible] by a [illegible] transformation formula the reliability coefficient [illegible] 1.00. This is the highest reliability coefficient [illegible] which the writer has found reported.[10]

There exemplify [illegible] problems:

(a) Fluctuation of the [illegible] reliability [illegible] cannot be determined from [illegible] data [illegible] manner.

(b) [illegible] the interaction [illegible] the [illegible] of the interaction [illegible] the rating about.

(c) The rating [illegible] [illegible] have to be analyzed [illegible] [illegible].

(d) Involves [illegible] [illegible] [illegible] of attitudes.

(e) Studies [illegible] [illegible] a [illegible] [illegible] [illegible]

current a work booklet on each of his officer subordinates. It is not difficult to believe that, in actual practice, the booklets are relegated to the senior's desk drawer to be pulled out from time to time and brought "up to date" by the process of recalling incidents to record. Although this practice undermines the plan to a considerable extent, it is probably better than the technique common in most rating plans in which the rater sits down once in three months or six months, or whatever the rating interval may be, and attempts a judgement based on what he can recall of his observations of the ratee during the interval since the last rating.

Other disadvantages of the system are:

(a) There is no provision for showing the ratee his rating.

(b) There is no provision for certifying to the ratee that his rating has been submitted to headquarters.

(c) The rating form is not designed to take advantage of modern machine methods of scoring, sorting, tabulation, and evaluation.

(d) The system does not provide for rating the raters.

(e) The system does not indicate the kind of experience with the ratee which the rater has used as a basis for his judgements. That is, one cannot tell from the rating whether the ratee was observed only in a school situation, an office situation, or in an operational position.

(f) The rating procedure and the form make no allowance

[illegible]

[illegible]

[illegible]

for the difference in duties and qualities required of officers depending on whether they are of the Line or the Staff, or whether they are specialists within those categories. It is possible, however, that the investigators, in choosing for the items on the rating scale in their work booklet "...incidents of how effective and ineffective officers had acted in particular military situations..."[14] have automatically accounted for differences in the duties and qualities required of various officers.

The Marine Corps Rating Form

Description and use. THE REPORT ON FITNESS OF OFFICERS OF THE UNITED STATES MARINE CORPS shown in Appendix H, is the briefest of the four military rating devices. The face of the form provides space for the usual biographical data and presents a few brief instructions to the rater.

Below these instructions is a graphic scale which lists the principal duties of a Marine officer in general terms with provisions for an evaluative rating on each. The possible judgements range from "Unsatisfactory" to "Outstanding".

Turning to the reverse side of the sheet we find a second graphic scale designed to measure certain attributes considered significant, including "Physical fitness" and "Intelligence". This scale also requires evaluative judgements ranging from "Unsatisfactory" to "Outstanding".

Following the second rating scale are several questions intended to elicit information as to the following:

for the difference in [illegible] of offices depending on whether [illegible] or the [illegible] or whether they are [illegible]. It is possible, however, that the [illegible] for the times [illegible] death of [illegible] particular military [illegible]. [illegible] assumed the [illegible] of various officers.

The Public Confidence [illegible]

[illegible] the [illegible] of the [illegible] the original [illegible] of the [illegible] and [illegible]

[illegible] the [illegible] [illegible] [illegible]

[illegible]

(a) Any adverse temperamental, moral, or physical characteristics which the ratee might have.

(b) Any other derogatory information which the rater may have on the ratee.

(c) The rater's attitude toward having the ratee under his command in wartime.

(d) The rater's estimate of the ratee's "General Value to the Service".

There is also space for any additional pertinent information or comment not covered elsewhere in the report.

Analysis. Recently the Headquarters of the Marine Corps began a project of research,[15]

> ...dealing with the revision of both officer and noncommissioned officer fitness reports with a view to improvement of those currently in use.

Proceeding with the portion of the project concerned with the officer fitness report, the investigators constructed six different types of rating devices based on sample rating forms and information obtained from the other three military services and from thirty leading commercial concerns. Using these six devices and the official Marine Corps form, comparable ratings were obtained on a population of 376 Marine officers of varying ranks with 208 other Marine officers of varying ranks acting as raters. The project plans call for correlation of these ratings with buddy ratings of the 376 subjects as a criterion.

(a) Any adverse temperamental, moral, or physical characteristics which the person might have.

(b) Any other derogatory information which the person may have on the record.

(c) The person's relative energy with the rules and his command in matters.

(d) The person's attitude of the remark "General also to the service".

There is also a [illegible] information or comment [illegible] to this report.

Analysis. Research at headquarters of the 34th Line Corps began a project of [illegible]

> ... [illegible] and nonpersonnel [illegible] reports with a view to improvement of their accuracy in use.

Proceeding [illegible] official [illegible] [illegible] [illegible]

The results, when they become available, will likely prove to be an interesting and valuable contribution to the store of knowledge in this field.

Evaluation. Until such time as the findings of the study described above become available, evaluation of the Marine Corps form must proceed largely on a basis of a consideration of its design.

The most noticeable advantage of the form is its brevity and the ease with which it can be filled out. This brevity is matched by the succinctness of the instructions to the rater. Other good points of the form are:

(a) It has the advantage inherent in a graphic scale in that the assessment of the ratee is itemized, showing his strong points and his weaknesses.

(b) It provides for a general overall evaluational statement by the rater.

(c) It provides for the report of pertinent information or comment not otherwise covered.

(d) It provides for an indication of the rater's attitude toward having the ratee under his command during war service.

Turning attention to the shortcomings of the form, we find that:

(a) The graphic scales are abbreviated and contain evaluative rather than descriptive terms to indicate the varying degrees of possession of a trait.

[illegible]

[illegible]

[illegible]

(b) The traits or qualities on the scales are not sufficiently well defined.

(c) There is no safeguard provided against the halo effect.

(d) One graphic scale includes "Physical fitness" and "Intelligence" among the qualities to be rated, a violation of the principle that qualities which can be more accurately determined by more objective means should not be listed on rating scales.

(e) The form is not designed to permit rapid and accurate scoring, tabulation, and evaluation of the data by modern machine methods.

(f) The form does not require that the rating be shown to the rates and there is no provision for assisting the rater in preparing for such an interview.

(g) The rating procedure does not provide for the rating being made by the superior with the most complete knowledge of the rates.

(h) The rating procedure and the form make no provision for the difference in duties and qualities required of officers depending on whether they are of the Line or the Staff, whether they are specialists within those categories, or whether they are in school or engaged in the duties of their respective classifications.

(i) The rating procedure makes no provision for either the averaging of the ratings made by several raters or the

review at higher echelons of ratings made at lower levels.

(j) Finally, the rating procedure makes no provision for certifying to the ratee that his rating has been submitted to headquarters.

SUMMARY

From the preceding description and discussion it is evident that the Navy's fitness report form is generally unsatisfactory. The Army's form is a superior one and is the most extensively tested of the four. Its principal disadvantage lies in the absence of any provision for acquainting the ratee with the details of his rating to aid him in self improvement. The form used by the Air Force appears to have considerable "face validity" and is an example of the best type of graphic rating scale. It possesses the added good feature of a provision for systematizing the rater's observation of the rater to a greater extent than is usual in rating systems. However, the Air Force system suffers from cumbersomeness in use. The chief advantage of the Marine Corps form is its brevity.

revolve at higher speeds or [illegible] morale.

(1) Finally, [illegible] for certifying to the [illegible] non-commissioned to headquarters.

From the preceding [illegible] it is evident that the [illegible] satisfactory. [illegible] most [illegible] advantage like [illegible] the rules will [illegible] improvement. [illegible] consideration [illegible] type of [illegible] features [illegible] tion [illegible] systems. [illegible] somewhat [illegible] in [illegible].

PART III

Chapter 1

CONCLUSIONS

From the foregoing chapters of description and discussion, certain conclusions may be drawn regarding the field of service rating in general and the design of an improved Naval officer's fitness report in particular. These conclusions may be summarized briefly as follows:

(a) Those in authority cannot eliminate the need for, nor escape the responsibility of assessing, either at intervals or more or less continuously, the value of their individual subordinates.

(b) In the present state of the science of psychological measurement, dependence must be placed on judgement in achieving these assessments.

(c) When assessing executives and Naval officers, the dependence on judgement is necessarily greatest due to the non-existence, to a large extent, of objective measures of executive performance.

(d) Human judgement is a complex product of many subtle, mutually interacting influences and, unless it is rather firmly founded on objective evidence, it is certain to be a variable entity, characterized by considerable inaccuracy.

(e) Lacking suitable means of obtaining objective evidence, we need some method of systematizing judgement and controlling or minimizing its weaknesses. A concomitant need,

which is relatively easy to fill, is provision of a device for recording judgements, thereby avoiding dependence on unreliable memory.

(f) Undismayed by the difficulty of the problem, investigators have sought energetically for a number of years (at least since the closing years of the nineteenth century), and are continuing the search, for a means of controlling or minimizing the vagaries of human judgement in service rating.

(g) Such means as have been devised for this purpose have been only partially successful. This is perhaps due to failure to filter out of the rating the reflection therein of the personality of the rater. It will be remembered that the rater marks the ratee not as the ratee is but as he perceives the ratee to be.

(h) It appears that further improvement of importance must await success in, (1) formulation and verification of objective criteria of executive performance and, (2) the development of a whole new theoretical basis, as suggested by Flanagan.[1] In the meantime, the best features of the existing systems should be used, with full realization of their limitations, since any tool, even an imperfect one, is better than none.

(i) Each of the several systems of service rating in common use today has its advantages and disadvantages.

(j) The most promising possibilities as types of a new Navy officer's fitness report form appear to be the forced

[illegible]

choice rating and the Probst check list device, although neither has been tested conclusively against a graphic rating scale of superior design. The forced choice form seems to be most resistant to rater's bias and halo while the check list as designed by Probst is more successful in achieving an approximately normal distribution of scores.

(k) There are certain extra features which should be a part of the rating system if it is to accomplish each of its objectives.

(l) A few of those features are incompatible in that inclusion of one of them will detract somewhat from the effectiveness of another. For example, it is desirable to omit from the rating form the key for scoring in order to deny the rater information with which he may control the overall mark he gives. (Such concealment is fundamental to the forced choice system.) But when the real meanings of the ratings are thus withheld, important information which both the rater and ratees should have is kept from them and the value of a post rating interview is greatly reduced.

With these considerations in mind, it is proper to advance to the problem of improving the rating system in use for evaluating Naval officer performance.

PART III

Chapter 2

RECOMMENDATIONS

First Recommendation

It has been concluded that the most promising possibilities as types of a new Navy officer's fitness report are the forced choice rating and the Probst check list device. A third, perhaps somewhat less encouraging, possibility is an improved graphic rating scale incorporating all the recommended features for increasing objectivity and resistance to halo and rater's bias.

As stated in the preceding chapter, neither the forced choice device nor the Probst check list is clearly the better in all respects, and neither has been tested conclusively against a graphic scale of superior design. It is recommended, therefore, that the Bureau of Naval Personnel experiment with a scale of each of the three types especially constructed for its use. The three devices should be tested simultaneously for reliability and for validity against improved criteria and using in both tests the largest practicable populations.

The improved criteria should be derived from a proper combination of buddy ratings and other pertinent information regarding the ratees such as the morale and attitudes of their subordinates, the effectiveness of the organizations which they headed, (as determined by the most objective means which can be devised), etc. The methods to be employed in isolating and

collecting this information and in achieving its proper combination are beyond the scope of this thesis and cannot be discussed here. Suffice it to say, however, that the importance of the criteria is so great that no practicable effort should be spared in carrying out this phase of the project.

Form

The administration of officer personnel in the Navy is such that, although in some cases, initiation of administrative action may take place in the field, final authority rests in the Bureau of Naval Personnel. A fitness report designed to give only information for such final action might very well be of the type which conceals the true nature of the rating made upon it from all who do not hold the scoring key.

However, as discussed in Chapter 1 of PART I, the efficiency evaluation has additional objectives. It will be remembered that supervisor improvement is one objective, improvement of the ratee is another, while aiding in the maintenance of morale and discipline completes the list. Maximum accomplishment of the first two of these will require that the rating be in some easily understood analytic form, whereas an overall score might be satisfactory for the last named purpose.

If either the check list or the improved graphic scale proves to be the most efficient, and consequently is adopted, the requirement for a display of the ratee's qualities mentioned above will be met. The forced choice form however, would not provide the desired display. Should the forced choice

collecting this information are to receive. [illegible] cumulation are beyond the scope of this thesis [illegible] discussed here. Suffice it to say, however, that [illegible] of the criteria is [illegible] should be spread in another, but this phase of the [illegible].

Fees

The [illegible] of office [illegible] such that, [illegible] motion [illegible] the Bureau of [illegible]. [illegible] give only [illegible] of the [illegible] upon it [illegible].

However, as [illegible] alone, [illegible] bered that [illegible] of the [illegible] morale [illegible] ment [illegible] in [illegible] score [illegible]

[illegible]

rating be adopted nevertheless, provision should be made for an additional rating, for purposes of analysis, on a graphic scale which the rater could show to the ratee and discuss with him.

Second Recommendation

If the first recommendation is not feasible, the less expensive (and also less conclusive) plan recently employed by the Marine Corps, and described herein, should be followed. The rating devices of the Army and the Air Force, suitably edited to eliminate obviously inapplicable items, should be tested with an improved graphic scale against improved criteria. The most effective device from this test group should become the new Navy officer's fitness report.

Third Recommendation

Even if it is impracticable to consider altering the basic features of the present rating system, it would be a distinct step forward to prepare and place in service an improved graphic scale. It would be essential that limited tests to determine reliability be performed but the requirement for validity must be met by giving the scale maximum face validity during design and construction.

Fourth Recommendation

No matter what the final decision may be regarding rating devices, a definite rater training program should be established as soon as practicable as a permanent part of the Navy's educational system. In this case, as for the introduction

[illegible] to adopted [illegible], [illegible] [illegible] [illegible] an additional [illegible], for [illegible] or analysis, on a [illegible] scale which [illegible] [illegible] [illegible] [illegible] with this.

[illegible]

If [illegible] [illegible], [illegible] [illegible] expensive (and also [illegible] [illegible] [illegible] [illegible] by the [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]. The [illegible] [illegible] [illegible] [illegible] [illegible]

of a new rating device, maximum benefit will accrue to the Service only if the initiation of the program is preceded by a careful promotional procedure designed to gain acceptance for the training plan.

The material to be taught in a rater training program has been discussed in an earlier chapter. Rater training courses offering this material should be included in the curricula of the U. S. Naval Academy, The U. S. Naval Schools (General Line) at Newport, R. I. and Monterrey, Calif., and the U. S. Naval ROTC units at civilian colleges and universities. Instructional pamphlets should be prepared and circulated to other Naval activities and, if funds permit, a training moving picture film should be produced. This film would be particularly effective in dramatizing the cycle of, (1) observation of the rates by the rater, (2) recollection by the rater, at rating time, of his observations, and (3) transformation of these recollections into marks on the rating blank. Other aspects of the problem such as the benefit to the Service, and the Nation, which will result from careful and conscientious marking and the harm to the Service, and the Nation, which will result from improper marking of the ratings could be emphasized in a film.

Subsidiary Recommendations

The designers of a rating form for Navy use should take cognizance of certain general considerations which arise from recognition of the principles of good service rating design and

of a new rating device, maximum benefits will accrue to the service only if the initiation of the program is preceded by a careful procedural [illegible] to gain acceptance for the training plan.

The material to be covered in a [illegible] has been discussed in an earlier section. Any training course effort, this material should be included in the curriculum of the U. S. Naval Academy [illegible]. Very [illegible] (General Line) or Aeronautical [illegible], [illegible] the U. S. Naval War College [illegible] [illegible] sion. Instructional [illegible] [illegible] lated to other Naval activities [illegible], a training moving picture film would [illegible] be particularly effective [illegible] (1) observation of the [illegible] (2) [illegible] rator, at [illegible], of [illegible] and (3) [illegible] action of [illegible] other [illegible] and the [illegible] material [illegible] [illegible] in a [illegible]

administration which were discussed in Chapter 2, PART II. The proposed actions regarding the more important of these factors are offered in the following paragraphs as subsidiary recommendations.

Information to be Elicited

The discussion in Chapter 3 of PART II described the difficulties attendant upon the determination of what leaders do, or what is more important, of what aspects of a leader's behavior are significant and are therefore objects of interest on the part of his superiors. The ideal efficiency evaluation device would be designed to elicit information on those significant behaviors, and scale items would be chosen accordingly. Selection of the items is a matter of judgement on an a priori basis and, apparently for some time to come, must remain so.

Authorities in this field are unanimous in the opinion that the items used on a rating scale should be chosen from among those composed by men who will later use the scale in rating their subordinates. This applies no matter whether the projected device will be a graphic scale, a check list, or a forced choice rating form. The assumptions upon which this view is based are, (1) the immediate superiors of the ratees are in the best position to know what behaviors are significant and, (2) having participated in its design, the raters will have more respect for the rating device and will use it more carefully. However, even though the specific items

administration which were discussed in Chapter 8, Part II. The proposed actions regarding [illegible] of these factors are offered in the following [illegible] educational recommendations.

Information to be Utilized

The [illegible] difficulties [illegible] do, or what [illegible] behavior are [illegible] best on the [illegible] motivation data [illegible] significant [illegible] length. [illegible] a [illegible] remedial [illegible]

Information [illegible] that the [illegible] access [illegible] reliable [illegible] the [illegible] or [illegible] this [illegible] retard [illegible] [illegible]

should be obtained from the suggestions of the raters, guidance may be furnished them by specification of the areas in which suggestions are desired.

Significant areas of information. The most important of the significant areas of information are, (a) leadership, (b) performance of duty, (c) special qualifications, and (d) appearance and general conduct. In addition there should be, (1) a list of standard situations in which the rater must indicate the extent to which he would place dependence on the ratee, (2) a specified difficult situation in which the rater should indicate his attitude toward having the ratee serve with him, and,(3) a space for a general evaluational statement and citation of any pertinent commendatory or derogatory material.

Taking a cue from Naval Leadership[1] we find that in the leadership area important items are loyalty up and down, moral and physical courage, honesty and fairness, common sense, good judgement, confidence, initiative, tact, self control, humor, modesty, enthusiasm, ability to delegate wisely, to establish communication up and down, to plan, to implement plans, to control operations, and knowledge of other principles of good organization and management.

There will be much overlap between the leadership and performance of duty areas but the latter should be included nevertheless. It will provide a means of marking officers who may happen to be serving in positions requiring a minimal dis-

should be obtained from the [illegible] of [illegible] guidance may be formulated than of specifications, the [illegible] which suggestions are [illegible].

Significant trends in [illegible]. The [illegible] of the significant trends of [illegible] are, (a) [illegible], (b) performance of [illegible], (c) [illegible], and (d) appearance and [illegible]. [illegible] should be, (f) a list of [illegible] indicate the extent to which [illegible] the [illegible] of the rates, (g) a specification [illegible] should indicate [illegible] with him, and (h) a [illegible] ment and direction [illegible] material.

Tool [illegible]. [illegible] the lateral [illegible] moral and [illegible], [illegible] good [illegible], [illegible] humor, [illegible] [illegible] to control [illegible] good [illegible].

[illegible]

[illegible]

play of leadership qualities. Initiative, ability to plan, to implement plans, to control operations, to delegate wisely are important qualities. Others are enthusiasm, perseverance, accuracy, thoroughness, conscientiousness, and resourcefulness.

The special qualifications section would be intended to record any special aptitudes or abilities which the officer might have developed but which probably would not be on record in the Bureau of Naval Personnel. Such things as adeptness at ship handling or organizing and training gun crews are important to record but are not so treated usually. The final area covers personal neatness and that reasonable attention to small details of regulations and customs which are one of the distinguishing marks of the excellent officer.

The superiority of the descriptive item over the evaluational has been pointed out. Because of this superiority, all items should be constructed to describe the behaviors which will indicate the presence or absence of the various qualities involved.

Specificity of Rating Device

Closely related to the subject of the significant areas of information to be covered, is the question of whether a single rating device can cover adequately all the important qualities which must be possessed by the officer corps of a modern navy. Included in this group of men, one will find in addition to the general line officer (executive), a long list

play of leadership qualities. Intuitively, ability to lead, to implement plans, to control operations, to motivate effectively are important qualities. Others are intelligence, persuasiveness, accuracy, observableness, consistiveness, and resourcefulness.

(4) The needed qualities to meet a special situation interested to record any special conditions or difficulties which the officer might have developed for action possibly would be an record in the process of final correspond. Such a list is essentially dependent on ability measure or evaluating certain traits and are important to record for use by the Personnel Officer. The final area covers personal qualities and that relate to attention to small details of procedure and [illegible] are one of the distinguishing [illegible] of [illegible] officers.

The subcategory of the psychiatric [illegible] emotional [illegible] Because of the [illegible] all items [illegible] which will [illegible] [illegible] [illegible]

of specialists in various technological fields, in communication, business and supply, law, the social sciences, and in medicine. In the sense that all Naval officers are executives, they must bring to their duties certain fundamental characteristics common to all branches. However, each of the specialties listed above requires also a different aggregation of peculiar knowledges and skills, and perhaps attitudes to some extent. But in considering the measurement of these special qualities, one finds that he is encroaching on the fields of intelligence testing, and aptitude and achievement testing where service rating has no place.

There is no satisfactory answer to this question. Authorities agree that a rating device should be designed for the specific group which it is intended should be rated by it and they advise that several different forms may be needed to rate the employees of a single commercial firm. Recommendations regarding the specificity of the executive rating form are not so unequivocal, however. This is not surprising when one recalls what has been said in Chapter 3 of PART II concerning the lack of information as to what aspects of an executive's performance are significant for the purposes of evaluation.

It is probable therefore, that a satisfactory answer will not be forthcoming until the findings of studies now in progress become available. These findings may reveal differences (as well as similarities) between the behaviors of general line officers and the various categories of specialists

of specialists in various technical fields, in communication, business and supply, law, the social sciences, and in medicine. In the same list all Naval officers are executives, they must bring to their duties certain fundamental characteristics common to all officers. In order, each of the specialties listed above requires also a different aggregation of particular knowledge and skills, and perhaps attitudes to some extent. But in considering the assessment of these special qualities, one finds that [illegible] on the fields of intelligence testing, and aptitude and achievement testing where service rating are no [illegible].

There is no satisfactory answer to this question. Authorities agree that a rating device should be established for the specific group which it is intended to [illegible] if it and they advise that several different [illegible] may be needed to rate the employees of a [illegible]. Nevertheless, regarding the specificity of the [illegible] are not so unequivocal. [illegible] This is at least [illegible] one reason why [illegible] the lack of information on the [illegible] performance and qualities [illegible] to [illegible].

It is [illegible] that a [illegible] will [illegible] progress [illegible] [illegible] areas [illegible] [illegible] [illegible]

of such magnitude that the design of special rating blanks for each will be seen as clearly necessary.

Three of the four military rating devices which have been examined make some attempt to secure a rating in the officer's specialty. This feature should be retained and expanded in scope in the new form, insofar as budgetary consideration will permit. Expansion might be accomplished by, (1) devising a basic rating form for all officers to be accompanied by special rating forms for certain classes of specialists, or (2) by giving more space on the single overall form to a description of and rating in a specialty.

If the plan for additional specialist rating blanks is adopted, the specialist officers might be grouped satisfactorily for rating purposes into, (1) technological and scientific specialists, (2) business and supply specialists, and (3) Medical and Dental officers.

Information to be Excluded

As previously discussed, qualities which may be determined by more objective means should not be measured by rating. Moreover, because of the importance of limiting the length and complexity of the form wherever possible, items which require entries that do not contribute information of value in reaching an assessment of the ratee should be omitted.

Length and Complexity

The question of the length of the rating form, the number of items on it, was discussed in Chapter 2 of PART II where

of such magnitude that a design of a social testing climate for each will be seen as clearly necessary.

Three of the four military testing services also have been examined with a view attempts to secure a testing in the officer's specialty. This feature should be emphasized and expanded in scope in the new Act, thereby an appropriate operation will result. Improvements might be accomplished by:

(1) devising a basic test [illegible] to be accompanied by special tests [illegible] elements of specialists, or (2) by [illegible] a single overall form to a description of [illegible] specialty.

If the Air Force [illegible] is adopted, the specialist [illegible] should be [illegible] for testing purposes [illegible] (1) psychological and aptitude specialists, (2) [illegible] specialists, and (3) medical and dental officers.

Information to be included

As previously discussed, testing [illegible] only a determined by [illegible] or [illegible] [illegible] Moreover, because of the importance of defining the length and complexity of the job [illegible], there which requires entries that do not [illegible] [illegible] or value in [illegible] ing an assessment of the [illegible] [illegible] systems.

Length and complexity

The [illegible] number of items [illegible] [illegible]

it was concluded that the longer forms are likely to be more reliable. However, as mentioned there, the cost and administrability cannot be ignored. To cover specifically each of the qualities mentioned earlier in this chapter would require a minimum of approximately thirty items. This number may seem excessive by the standards of those, who, on the basis of the results of factor analysis, feel that rating scales measure only one or two qualities at most. But, judging from the experience of Probst, thirty items may not be too many. He has reported that the average number of items checked on his scale, which though it lists 101 items specifies that only those which are applicable to the rates are to be checked, is between twenty and twenty-five.[2]

It would be desirable to limit the form to only the two sides of a single sheet of paper 8½ inches wide and 14¾ inches long, the size of the present Navy fitness report form. The requirement for carbon copies should be eliminated, insofar as possible, as it complicates preparation and filing and increases costs. There appears to be no need for copies because "Incoming (fitness) reports are filed in a jacket of their own. This is available for all personnel actions,..."[3]

So long as the Navy Regulations provide that prompt submission of fitness reports is a joint responsibility of the rates and rater, some means of notification to the rates that the transaction has been accomplished is desirable. The present Navy fitness report form provides a carbon copy of the

it was concluded that [illegible] are likely to be more reliable. Moreover, [illegible] trustworthiness [illegible] the qualities [illegible] a minimum of [illegible] seem excessive [illegible] of the [illegible] manners only [illegible] the experience of [illegible] He has [illegible] his souls, [illegible] those which are [illegible] between [illegible]

It would be [illegible] two sides of a [illegible] Knolsa [illegible] The requirement [illegible] far as possible, [illegible] innocence [illegible] cause "[illegible] their own. [illegible]

[illegible] subjects [illegible] the [illegible] that [illegible] pro [illegible]

biographical section of the report for the ratee. This feature or its equivalent should be retained. Indeed, there is much to be said for furnishing a complete copy of the entire rating to the ratee. With a permanent and complete record of his superiors' judgements of him in his possession, he would be in a much better position to plan action toward self improvement.

Machine Processing

The most efficient method for scoring, sorting, or tabulating a large number of similar reports is the machine method discussed briefly in Chapter 2, PART II. The facility and accuracy with which these operations can be accomplished by machinery render the dependence on older processes unwise.

There has been some experimentation in the Bureau of Naval Personnel with a revision of the present fitness report form "designed to facilitate punching it on an IBM card".[4] Although the revised form was not adopted, any new design of fitness report definitely should include provisions for machine processing in the Bureau. The Navy cannot afford to ignore the important aid to advanced personnel practices which the technique of machine processing of fitness reports affords. It is important in the unprecedentedly large peacetime Navy of today and would be vastly more important when the Navy is required to expand to meet a national emergency.

REFERENCES

References are arranged numerically by chapters.

PART I

Chapter 1

[1]M. N. Hay, "Selling a Rating Plan," Personnel, 18:42, 1941.

[2]M. J. Jucius, Personnel Management, (Chicago: Richard D. Irwin, 1947), p. 426.

[3]J. Tiffin, Industrial Psychology, Second Edition, (New York: Prentice-Hall, Inc., 1947), p. 321.

[4]S. D. Marble, "A Performance Basis for Employee Evaluation," Personnel, 18: 218, 1942.

[5]L. D. White, Introduction to the Study of Public Administration, (New York: The Macmillan Company, 1948), P. 420

[6]Tiffin, op. cit., p. 321.

[7]Jucius, op. cit., P. 427.

[8]D. Yoder, Personnel Management and Industrial Relations, (New York: Prentice-Hall, Inc., 1948), p. 337.

[9]Tiffin, op. cit., pp. 330-332.

[10]P. Pigors and C. A. Myers, Personnel Administration a Point of View and a Method, First Edition, (New York: McGraw-Hill Book Company, Inc., 1947), p. 171.

[11]W. E. Mosher and J. D. Kingsley, Public Personnel Administration, Revised Edition, (New York: Harper & Brothers, 1941), p. 481.

[12]Tiffin, op. cit., p. 331.

[13]Jucius, op. cit., p. 427.

[14]Register of Commissioned and Warrant Officers of the United States Navy and Marine Corps, January 1, 1949. (Washington: Government Printing Office).

[15]United States Navy Regulations, 1948, (Washington: Government Printing Office).

Bibliography

References are arranged alphabetically by surname.

Part I

Chapter I

[illegible] 1941.

[illegible] D. Irwin, 1947), [illegible].

[illegible] (New York: [illegible], 1947), p. 361.

[illegible] vation," [illegible], 1962.

[illegible] misinterpretation, (New York: [illegible], 1962), p. 300

[illegible] Griffin, op. cit., p. 301.

[illegible] Jacobs, op. cit., p. 197.

[illegible] Yoder, Personnel Management and Industrial Relations, (New York: [illegible], [illegible]).

[illegible] Yoder, op. cit., [illegible].

[illegible] a Point of View [illegible] Behavior, [illegible] (New York: [illegible] Hill Book [illegible], 1957).

[illegible] Administration, revised [illegible] (New [illegible] 1941), [illegible].

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

REFERENCES

PART I

Chapter 2

No References

PART II

Chapter 1

1 D. Yoder, Personnel Management and Industrial Relations, (New York: Prentice-Hall, Inc., 1948), p. 353.

2 Naval Leadership, First Edition, (Annapolis, Maryland: U. S. Naval Institute, 1949), p. 84.

3 R. Stagner, Psychology of Personality, Second Edition, (New York: McGraw-Hill Book Company, Inc., 1948), p. 93.

4 G. W. Allport and H. S. Odbert, "Trait Names: A Psycholexical Study", Psychological Monographs, 47: vi, 1936.

5 D. Krech and R. S. Crutchfield, Theory and Problems of Social Psychology, First Edition, (New York: McGraw-Hill Book Company, Inc., 1948), p. 77.

6 Ibid., p. 94.

7 H. A. Murray, "The Effect of Fear upon Estimates of Maliciousness of Other Personalities", Journal of Social Psychology, 4:310-329, 1933.

8 C. Leuba and C. Lucas, "The Effects of Attitudes on Descriptions of Pictures", Journal of Experimental Psychology, 35: 517-524, 1945.

9 F. Heider, "Social Perception and Phenomenal Causality", Psychological Review, 51: 358-363, 1944.

10 . Zillig, "Einstellung and Aussage", Z. Psycho., 106: 67-106, 1928, cited by Krech and Crutchfield, op. cit., p. 105.

11 Krech and Crutchfield, op. cit., pp. 150-151.

REFERENCES

PART II

Chapter 1

[12] An attitude is similar to but stronger than a belief, being more emotionally loaded and enduring. Krech and Crutchfield make this distinction, op. cit., pp. 151-153.

[13] Ibid., p. 125 and p. 188.

[14] F. C. Bartlett, "Social Factors in Recall", T. M. Newcomb, E. L. Hartley, and others, Readings in Social Psycology, (New York: Henry Holt and Company, 1947), p. 72.

[15] J. M. Levine and G. Murphy, "The Learning and Forgetting of Controversial Material", Newcomb and Hartley, op. cit., pp. 108-115.

[16] Krech and Crutchfield, op. cit., p. 191.

[17] Ibid., p. 194.

[18] H. E. Burtt, Principles of Employment Psychology, Revised Edition, (New York: Harper & Brothers, 1942), p. 392.

[19] L. Stockford and H. V. Bissell, "Factors Involved in Establishing a Merit Rating Scale", Personnel, 26: 85-116, 1949. In this article the authors describe extensive experiments incident to the design and construction of a merit rating device. Their investigations involved, among other things, ratings and reratings by 77 supervisors.

[20] Ibid., p. 105

[21] H. L. Hollingworth, Judging Human Character, (New York: D. Appleton and Company, 1928), p. 76.

[22] Stagner, op. cit., p. 140.

[23] W. L. Sisson, "Forced Choice--The New Army Rating System", Personnel Psychology, 3: 376, 1948.

[24] L. J. Cronbach, Essentials of Psychological Testing, (New York: Harper & Brothers, 1949), p. 388.

[25] J. P. Guilford, Psychometric Methods, First Edition, (New York: McGraw-Hill Book Company, Inc., 1936), p. 272.

PART II

Chapter I

12 [illegible] in stylish [illegible] Chap. 2 [illegible] Hiel, being [illegible] Hiel. Leaded [illegible] [illegible] Cuneiform [illegible] [illegible], op. cit., p. 121-122.

13 Ibid., p. 122 and p. 123.

14 F. C. Avisiri, [illegible] (1917), I. [illegible] However, F. H. Bertling, and others, [illegible] *[illegible]* (New York: Kelly [illegible] Co., 1947), p. [illegible].

15 [illegible] *[illegible]* op. cit., p. 124-125.

16 [illegible] op. cit., p. 124.

17 Ibid., p. 124.

18 [illegible] Burke, [illegible] Revised [illegible] (New York: [illegible] 1962), p. 302.

19 [illegible] Lewis, "Master Revised in [illegible] *[illegible]*, XIV, 32-112, 1939. [illegible] [illegible] [illegible]

REFERENCES

PART II

Chapter 1

[26] J. E. Greene, Measurements of Human Behavior, (New York: The Odyssey Press, 1941), p. 707.

[27] Stockford and Bissell, op. cit., p. 105.

[28] W. V. Bingham, "Halo Valid and Invalid", Journal of Applied Psychology, 23: 228, 1939.

[29] E. B. Hurlock, "An Evaluation of Certain Incentives Used in School Work", Journal of Educational Psychology, 16: 157, 1925.

[30] Naval Leadership, op. cit., p. 73.

Chapter 2

[1] P. M. Symonds, "On The Loss of Reliability in Ratings Due to Coarseness of the Scale", Journal of Experimental Psychology, 7: 456-461, 1924.

[2] Ibid., p. 460.

[3] J. P. Guilford, Psychometric Methods, First Edition, (New York: McGraw-Hill Book Company, Inc., 1936), p. 268.

[4] E. E. Ghiselli and C. W. Brown, Personnel and Industrial Psychology, First Edition, (New York: McGraw-Hill Book Company, Inc., 1948), p. 107.

[5] M. J. Jucius, Personnel Management, (Chicago: Richard D. Irwin, 1947), p. 448 shows a profile derived from a rating.

[6] H. E. Burtt, Principles of Employment Psychology, Revised Edition, (New York: Harper & Brothers, 1942), pp. 373-374.

[7] D. H. Sisson, "Forced Choice--The New Army Rating System," Personnel Psychology, 3: 371-373, 1948.

[8] J. Tiffin, "Merit Rating: Its Validity and Techniques", American Management Association Personnel Series, Number 100, (New York: American Management Association, 1948), p. 19.

APPENDIX

Part II

Chapter I

26 [illegible], Reproduction of [illegible] Artificial [illegible] (New York: [illegible] Press, 1953), p. 107.

27 [illegible], op. cit., p. 108.

28 [illegible], "Film [illegible] Material", Journal of Applied Psychology, 33: [illegible], 1949.

29 [illegible], "An [illegible] of Certain Innovative Uses in Formal [illegible]", Journal of Industrial Psychology, 13: 157, 1952.

30 Tape Recorders, op. cit., p. 72.

Chapter II

1 [illegible], "[illegible] Use of Tape Recorders at [illegible]", Journal of Experimental Psychology, 71: 460-461, 1954.

2 Ibid., p. 460.

3 [illegible], Electronic Data [illegible], First Edition, (New York: McGraw-Hill Book Company, Inc., 1955), p. 288.

4 [illegible], Clinical [illegible] Trial Preparation, (New York: Columbia-Princeton Company, Inc., 1955), p. 157.

5 [illegible], Instructional [illegible], (New York: [illegible], 1957), p. [illegible].

6 [illegible], Principles of [illegible] Revised Edition, (New [illegible], 1957), pp. 270-[illegible].

7 [illegible], op. cit., p. [illegible].

8 [illegible], op. cit., p. [illegible].

REFERENCES

PART II

Chapter 2

[9] Burtt, op. cit., p. 182.

[10] S. D. Marble, "A Performance Basis for Employee Evaluation", Personnel, 18: 217-226, 1942.

[11] E. B. Knauft, "Construction and Use of Weighted Check-list Rating Scales for Two Industrial Situations", Journal of Applied Psychology, 32: 63-70 1948.

[12] Burtt, op. cit., p. 354.

[13] J. Tiffin and W. Musser, "Weighting Merit Rating Items", Journal of Applied Psychology, 26: 575, 1942.

[14] J. Tiffin, Industrial Psychology, Second Edition, (New York: Prentice-Hall, Inc., 1947), pp. 501-505.

[15] Tiffin and Musser, op. cit., p. 583. Also see J. Tiffin, loc. cit.

[16] Guilford, op. cit., p. 272.

[17] E. L. Thorndike, Individuality, (Boston: Houghton Mifflin Company, 1911), p. 13.

[18] D. C. Adkins and others, Construction and Analysis of Achievement Tests, (Washington: Government Printing Office, 1947), pp. 98-103.

[19] W. E. Mosher and J. D. Kingsley, Public Personnel Administration, Revised Edition, (New York: Harper & Brothers 1941), pp. 494-495.

[20] Jucius, op. cit., pp. 435 and 440.

[21] Burtt, op. cit., p. 369.

[22] Guilford, op. cit., p. 278.

[23] L. Stockford and H. W. Bissell, "Factors Involved In Establishing a Merit Rating Scale", Personnel, 26: 105, 1949.

[24] Marble, op. cit., p. 220.

BIBLIOGRAPHY

Part II

Chapter 2

9 Burtt, op. cit., p. 320.

10 G. E. Martin, "A Instrument for Judgment Evaluation", Personnel, 20: 217-220, 1942.

11 E. J. Cowles, "Construction [illegible] Rating Scales for [illegible] Personnel Work", Journal of Applied Psychology, 26: 99-70 1942.

12 Burtt, op. cit., p. 324.

13 J. Tiffin and [illegible] Lessner, "Validating Merit Rating Items", Journal of Applied Psychology, 26: 279, 1942.

14 J. Tiffin, Industrial Psychology, Second Edition, (New York: Prentice-Hall, Inc., 1947), pp. 201-205.

15 Tiffin [illegible] op. cit., p. 325. [illegible] J. Tiffin, loc. cit.

16 Guilford, op. cit., p. 272.

17 A. L. Wisenblke, Multivariability, (Boston: Houghton Mifflin Company, 1951), p. 15.

18 H. C. Adkins and others, Construction and Analysis of Achievement Tests, (Washington: Navy Department, Bureau of Aeronautics, 1947), pp. 98-100.

19 W. [illegible] Cooper and [illegible], Personnel Administration, (Ginn and Company, [illegible] Boston, 1941), pp. 154-156.

20 Jenkins, op. cit., p. 401 - 440.

21 Burtt, op. cit., p. 326.

22 Guilford, op. cit., p. 270.

23 L. [illegible] and [illegible], "Merit [illegible] In Particular [illegible] Rating Scale Data", Personnel, 20: 100, 1940.

24 Burtt, op. cit., p. [illegible].

REFERENCES

PART II

Chapter 2

[25] L. J. Cronbach, Essentials of Psychological Testing, (New York: Harper & Brothers, 1949), p. 401.

[26] Adkins, op. cit., p. 169.

[27] J. Tiffin, "Merit Rating: Its Validity and Techniques", American Management Association Personnel Series, Number 100, (New York: American Management Association, 1946), p. 21.

[28] C. H. Lawshe, H. C. Kephart, and E. J. McCormick, "The Paired Comparison Method for Rating Performance of Industrial Employees", Journal of Applied Psychology, 33: 69-77, 1949.

[29] L. L. Thurstone, "The Method of Paired Comparisons for Social Values", Journal of Abnormal and Social Psychology, 21: 384-400, 1927.

[30] Guilford, op. cit., pp. 217-243.

[31] Ibid., pp. 223 and 235.

[32] Lawshe, Kephart, and McCormick, op. cit., p. 76.

[33] M. B. Greene, Measurements of Human Behavior, (New York: Odyssey Press, 1941), p. 692.

[34] This discussion of the Man to Man scale is condensed from the excellent treatment of the subject by W. D. Scott, R. C. Clothier, S. B. Mathewson, and W. R. Spriegel, Personnel Management, Third Edition, (New York: McGraw-Hill Book Company, Inc., 1941), pp. 215-224.

[35] An organization established by a group of thirty important business firms in 1916, see Scott, Clothier, Mathewson, and Spriegel, Ibid., p. 214.

[36] J. B. Probst, Service Ratings, (Baltimore: Lord Baltimore Press, 1931), pp. 1-94.

[37] Ibid., p. 80.

[38] Ibid., pp. 60, 71, and 76.

[39] Ibid., p. 51.

NOTES

PART II

Chapter 5

[illegible] E. [illegible], [illegible] [illegible] (New York: [illegible], [illegible]), [illegible].

[illegible], op. cit., p. [illegible].

[illegible], "[illegible] [illegible] [illegible] [illegible]," [illegible] (New York: [illegible], [illegible]), p. [illegible].

[illegible], "The [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]," [illegible], [illegible], [illegible].

[illegible], "[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [

REFERENCES

PART II

Chapter 2

40 Ibid., p. 68.

41 Ibid., p. 54.

42 Ibid., p. 53.

43 Knauft, op. cit., p. 65.

44 Ibid., p. 69.

45 L. D. White, Introduction to the Study of Public Administration, (New York: The Macmillan Company, 1948) p. 415.

46 Mosher and Kingsley, op. cit., pp. 492-496.

47 J. B. Probst, Measuring and Rating Employee Value, (New York: The Ronald Press Company, 1947), p. 70.

48 Guilford, op. cit., p. 279.

49 F. F. Bradshaw, The American Council on Education Rating Scale, Its Reliability, Validity, and Use, (New York: Archives of Psychology, Number 119, 1930) p. 58.

50 A. V. MacCullough, "Periodic Review of Management Personnel", Personnel, 24: 140-149, 1948.

51 D. Yoder, Personnel Management and Industrial Relations, (New York: Prentice-Hall, Inc., 1946), p. 353.

52 Tiffin, "Merit Rating: Its Validity and Techniques", op. cit., p. 19.

53 Burtt, op. cit., p. 396.

54 Ibid., p. 441.

55 For a report of other experimental evidence showing a similar result, see Ghiselli and Brown, op. cit., p. 109.

56 Gisson, op. cit.,

57 Ibid.

CONTINUED

PART II

Chapter 3

40 Ibid., p. [illegible].

41 Ibid., p. 34.

42 Ibid., p. 32.

43 Kandel, op. cit., p. [illegible].

44 Ibid., p. [illegible].

45 L. [illegible], Introduction to the [illegible] of Public Administration, (New York: The Macmillan Company, 1950) p. [illegible].

46 Gaston and Blanchard, op. cit., [illegible].

47 J. D. Porter, Governing the [illegible] Metropolitan Dallas, (New York: The Ronald Press Company, 1957), [illegible].

48 Gulliver, op. cit., [illegible].

49 F. J. Goodnow, [illegible], Politics and Administration (New York: The Macmillan Company, 1900), [illegible].

50 A. [illegible] Greenfield, "Functional Origins of Our Civil Personnel", Indianapolis, IN: Indiana, 1965.

51 F. [illegible], Personnel Management in Municipal Governments (New York: Ronald Press Company, 1958), p. [illegible].

52 Griffin, Merit [illegible], op. cit., p. [illegible].

53 [illegible], op. cit., [illegible].

54 Ibid., p. 412.

55 [illegible] a short [illegible] a similar [illegible] op. cit., [illegible].

56 [illegible], op. cit.,

57 Ibid.

REFERENCES

PART II

Chapter 2

58 L. L. Thurstone, The Reliability and Validity of Tests, (Ann Arbor: Edwards Brothers, 1933), pp. 33-39, cited in Adkins, op. cit. p. 156.

59 Thurstone, op. cit., pp. 46-51, cited in Adkins, op. cit., p. 166.

60 C. H. Lawshe, Principles of Personnel Testing, First Edition, (New York: McGraw-Hill Book Company, Inc., 1948), p. 28.

61 Stockford and Bissell, op. cit., p. 103.

62 E. Ewart, S. L. Seashore, and J. Tiffin, "A Factor Analysis of an Industrial Merit Rating Scale", Journal of Applied Psychology, 25: 485, 1941.

63 W. V. Bingham, "Halo Valid and Invalid", Journal of Applied Psychology, 23: 226, 1939.

64 J. B. Probst, Measuring and Rating Employee Value (New York: The Ronald Press Company, 1947), p. 20.

65 Ewart, Seashore, and Tiffin, loc. cit.

66 G. D. Halsey, Making and Using Industrial Service Ratings, First Edition, (New York: Harper & Brothers, 1944), pp. 12-13.

67 Yoder, op. cit., pp. 336-337.

68 Probst, op. cit., pp. 4-5, p. 29.

69 W. D. Scott, R. C. Clothier, S. B. Mathewson, and W. R. Spriegel, Personnel Management, Third Edition, (New York: McGraw-Hill Book Co., Inc., 1941), p. 234.

70 T. G. Armstrong, "Talking Your Ratings", Personnel, 26: 112-113, 1943.

71 Bisson, op. cit.

[illegible]

[illegible] II

Chapter 8

58 L. L. Thurstone, *The Reliability and Validity of Tests*, (Ann Arbor: Edwards Brothers, 1932), pp. 58-62, cited in Adkins, *op. cit.*, p. 116.

59 Thurstone, *op. cit.*, pp. 16-21, cited in [illegible], *op. cit.*, p. 166.

60 C. H. Lawshe, *Principles of Personnel Testing*, First Edition, (New York: McGraw-Hill Book Co., Inc., 1948), p. 28.

61 Lawshe and [illegible], *op. cit.*, p. 108.

62 M. Ward, J. [illegible], and [illegible], "Labor Analysis of an Industrial Task with [illegible]," *Journal of Applied Psychology*, 28: 488, 1944.

63 [illegible], "Data Tells [illegible]," *Journal of Applied Psychology*, 28: 488, 1944.

64 J. [illegible] Fryer, *Examining and Training Employee Talent* (New York: The Ronald Press Company, 1957), p. 10.

65 [illegible], Sanborn, and Tiffin, *Loc. cit.*

66 J. [illegible] Tiffin, *Testing and Labor Productivity Studies Analysis*, Silver Edition, (New York: Harper & Brothers, 1943), pp. 12-13.

67 Yoder, *op. cit.*, pp. 320-327.

68 Bennet, *op. cit.*, pp. [illegible], p. 40.

69 [illegible] Scott, J. C. Clothier, [illegible], and W. R. Spriegel, *Personnel Management*, 5th Edition, (New York: McGraw-Hill Book Co., Inc., 1961), p. 324.

70 [illegible], "Tailor Made [illegible]," *[illegible]*, 26: 112-113, 1942.

71 Lawson, *op. cit.*

REFERENCES

PART II

Chapter 2

72 R. J. Wherry and D. H. Fryer, "Buddy Ratings: Popularity Contest or Leadership Criteria", Personnel Psychology, 2: 147-159, 1949.

73 R. H. Bittner, "Developing an Industrial Merit Rating Procedure", Personnel Psychology, 1: 409, 1948.

74 Halsey, op. cit., p. 119.

75 Probst, op. cit., p. 43.

76 J. B. Probst, Service Ratings, (Baltimore: Lord Baltimore Press, 1931), p. 34.

77 Yoder, op. cit., p. 359.

78 L. D. White, Introduction to the Study of Public Administration, Third Edition, (New York: The Macmillan Co., 1948), p. 419.

79 Ibid.

80 R. Bittner, "Developing an Employee Merit Rating Procedure", Personnel, 25: 283, 1949.

81 A. S. Knowles, "Merit Rating and Labor Management", Personnel, 17: 33, 1940.

A. S. Knowles, "Merit Rating of Supervisors Foremen and Department Heads", Personnel, 17: 189, 1940.

82 J. Tiffin, "Merit Rating: Its Validity and Techniques", American Management Association Personnel Series, Number 100, (New York: American Management Association, 1948), p. 16.

83 R. S. Driver, "Training as a Means of Improving Employee Performance Rating", Personnel, 18: 369, 1942.

84 R. H. Bittner, "Developing an Industrial Merit Rating Procedure", Personnel Psychology, 1: 421-422, 1948.

85 Driver, loc. cit.

BIBLIOGRAPHY

PART II

Chapter 2

72 R. J. Harvey, et al., op. cit., "Family Business Succession Survey Contest or Fundamental Research", Personnel Psychology, 27: 147-159, 1974.

73 R. E. Fiedler, "Analysis of Industrial Training Procedure", Personnel Psychology, 11: 90, 1958.

74 Mahoney, op. cit., p. 134.

75 Bassett, op. cit., p. 55.

76 J. J. Corr, Training Manual, (Northbrook: Dartnell Corp. Publications Group, 1971), p. 76.

77 Yoder, op. cit., p. 224.

78 L. Urwick, Introduction to the Theory of Public Administration, (New York: McGraw Hill, New York Book Co., 1949), p. 42.

79 Ibid.

80 R. Litter, "Periodic Employee Performance Rating Procedure", Personnel, 50: 555, 1973.

81 R. Cowles, "Merit Rating and Labor Management", Personnel, 17: 32, 1940.

82 L. Lawrence, "Merit Rating of Supervisors Versus and Patrolmen Essay", Personnel, 17: 134, 1940.

83 J. Zirlin, "Merit Ratings are Valuable for Promotion", American Journal of Nursing Personnel Review, Vol. 1, (New York: Publication Department, Nursing Service, 1951).

84 R. E. Dreyer, "Problems of Human Resources Employee Information", Personnel, 1: 1-11, 1974.

85 R. N. Lyons, "Problems in Performance Appraisal Procedure", Personnel Journal, 53: 481-484, 1974.

86 Pigors, Leon, op. cit.

REFERENCES

PART II

Chapter 2

[86] Stockford and Bissell, op. cit., pp. 115-116.

[87] The forced choice rating was only briefly mentioned in this chapter. A more complete discription and evaluation of a particular rating device of the forced choice type will be found in Chapter 4 of PART II.

[88] Yoder, op. cit., pp. 365-369.

Chapter 3

[1] D. Yoder, Personnel Management and Industrial Relations, (New York: Prentice-Hall, Inc., 1946), p. 389.

[2] H. E. Burtt, Principles of Employment Psychology, Revised Edition, (New York: Harper & Brothers, 1942), p. 558.

[3] J. B. Probst, Measuring and Rating Employee Value, (New York: The Ronald Press Company, 1947), pp. 73-74.

[4] L. Stockford and H. W. Bissell, "Factors Involved in Establishing a Merit Rating Scale", Personnel, 26: 108-109, 1949.

[5] M. W. Richardson and G. F. Kuder, "Making a Rating Scale that Measures", The Personnel Journal, 12: 36-39, 1933.

[6] C. H. Lawshe, N. C. Kephart, and A. J. McCormick, "The Paired Comparison Method for Rating Performance of Industrial Employees", Journal of Applied Psychology, 33: 69-77, 1949.

[7] Reported in H. L. Bills, "Merit Rating and Seniority", American Management Association Personnel Series, Number 127, (New York: American Management Association, 1949), p. 39.

[8] C. H. Lawshe, Principles of Personnel Testing, First Edition, (New York: McGraw-Hill Book Company, Inc., 1948) p. 176.

[9] L. J. Cronbach, Essentials of Psychological Testing, (New York: Harper & Brothers, 1949), p. 401.

BIBLIOGRAPHY

PART II

Chapter 3

86 Goodchild and Lancaster, op. cit., p. 7-12.

87 The forces chosen [illegible] only briefly mentioned in this chapter. [illegible] of a particular [illegible] be found in chapter 4 on [illegible].

88 Yoder, op. cit., pp. 38-[illegible].

Chapter 2

1 [illegible] Yoder, Personnel and [illegible] Relations. (New York: Prentice-Hall, Inc., [illegible]).

2 [illegible] Dorsey, [illegible] Revised Edition. (New York: Harper [illegible], 1948), p. 212.

3 [illegible] (New York: The [illegible]).

4 [illegible] in Establishing a Merit System, [illegible] 1940.

5 [illegible] Scale Clerk Examiner, [illegible] 1953.

6 [illegible] "The Initial [illegible] Journal of [illegible] Quarterly, [illegible] 1940.

7 Reported [illegible] American Management Association Personnel Series [illegible] (New York: American [illegible]).

8 [illegible] Edition, (New York: [illegible]).

9 [illegible] (New York: [illegible]).

REFERENCES

PART II

Chapter 3

10 Probst, op. cit.

11 M. D. Knauft, "Construction and Use of Weighted Check List Rating Scales for Two Industrial Situations", Journal of Applied Psychology, 32: p. 65 and p. 69.

12 Development of the sociometric method is attributed to J. L. Moreno by R. Stagner, Psychology of Personality, Second Edition, (New York: McGraw-Hill Book Company, Inc., 1948), p. 28. Stagner refers to J. L. Moreno, "Who Shall Survive?" Nervous Mental Disability Monograph Series, Number 58.

13 R. J. Wherry and D. H. Fryer, "Buddy Ratings: Popularity Contest or Leadership Criteria", Personnel Psychology, 2: 147-159, 1949.

14 D. E. Hisson, "Forced Choice--The Net Army Rating System", Personnel Psychology, 1: 368-369, 1948.

15 Wherry and Fryer, op. cit.

16 Methodological Investigation of the Forced Choice Technique Utilizing the Officer Description and the Officer Evaluation Blanks, Study #701, PR 4061-04, (Washington: Department of the Army, 6 July, 1948), p. 5.

17 H. M. Littner, "Developing an Industrial Merit Rating Procedure", Personnel Psychology, 1: 410, 1948.

18 Ibid., p. 409.

19 A letter from Dr. E. G. Brundage, Head, Classification and Field Research Branch, Bureau of Naval Personnel, Department of the Navy, Washington, D.C., File No. Pers-152-af, Serial H819, 6 October, 1949.

20 Selection boards are committees of senior officers organized from time to time to consider officers eligible, by reason of seniority and experience, for promotion and to select from those considered, those who are to be promoted.

REFERENCES

PART II

Chapter 3

[21] Flag rank is the term applied to ranks in the Navy above that of Captain.

[22] See C. L. Shartle, Leadership and Executive Performance, Personnel, 25: 370-380, 1949.

[23] Personnel Research Board, The Ohio State University, Studies in Naval Leadership, Technical Report, (Columbus, Ohio: The Ohio State University Research Foundation, June 15, 1949), p. 8.

Chapter 4

[1] A letter from Dr. W. G. Brundage, Head, Classification and Field Research Branch, Bureau of Naval Personnel, Department of the Navy, Washington, D.C., File No. Pers-152-slf, Serial N949, 5 April, 1950.

[2] An Evaluation of the Report on the Fitness of Officers, P-4114-dmc, Project No. 427, (Washington: Training, Standards and Curriculum Division, Bureau of Naval Personnel, Department of the Navy, July, 1945), p. 1.

[3] Lieutenant Commander M. . Gagle, "The Navy's Fitness Report System", United States Naval Institute Proceedings, 73: 730, July, 1947.

[4] Construction of Three Measures for Instructor Evaluation, Report 383-1-5, RBH Project 128, (New York: Richardson, Bellows, Henry, And Company, Inc., July 30, 1945), p. 15.

[5] L. L. Thurson, "Forced Choice--The New Army Rating System", Personnel Psychology, 1: 365-381, Autumn, 1948.

[6] Ibid., pp. 369-380.

[7] Ibid., p. 369.

[8] Ibid., p. 374.

[illegible]

[illegible] II

Chapter 3

21 [illegible] in the [illegible] to focus [illegible] the navy above the [illegible].

22 [illegible] S. [illegible], [illegible] [illegible] [illegible] [illegible] **Radio**, [illegible], 201 [illegible], 1944.

23 [illegible] **Studies in Naval [illegible]**, [illegible] **The Data [illegible]**, [illegible], p. 2.

Chapter 4

1 [illegible] [illegible] [illegible] [illegible]

REFERENCES

PART II

Chapter 4

[9] Ibid., p. 367.

[10] The 1948 Standardization of Officer Efficiency Report, WD AGO FORM 67-1, WD AGO PRT Report 799, (Washington: Adjutant General's Office, Department of the Army), 7 June, 1949.

[11] Bisson, op. cit., p. 375.

[12] H. O. Preston, The Development of a Procedure for Evaluating Officers in the United States Air Force, (Pittsburgh: American Institute for Research, 7 July, 1948), pp. 1-67.

[13] Ibid., Technical Appendices and Notes, p. F-17.

[14] Preston, op. cit., p. 66.

[15] A Memorandum issued by the Procedures Analysis Office of the Office of the Director of Personnel, Headquarters U. S. Marine Corps, serial 2295-70 PA-1430-mtp, dated November 10, 1949, p. 1.

PART III

Chapter 1

[1] J. C. Flanagan, "A New Approach to Evaluating Personnel," Personnel, 26: 37, 1949.

Chapter 2

[1] Naval Leadership, First Edition, (Annapolis, Maryland; U. S. Naval Institute, 1949), pp. 135-245.

[2] J. B. Probst, Measuring and Rating Employee Value, (New York: The Ronald Press Company, 1947), p. 37.

REFERENCES

PART III

Chapter 2

[3] From a letter written by Dr. L. G. Brundage, Head, Classification and Field Research Branch, Bureau of Naval Personnel, Department of the Navy, Washington, D.C., File No. Pers-152-slf, Serial H949, 5 April, 1950.

[4] Ibid.

BIBLIOGRAPHY

A. BOOKS

Adkins, D. C., and others, Construction and Analysis of Achievement Tests. Washington: Government Printing Office, 1947. 202 pp.

Bartlett, F. C., "Social Factors in Recall", pp. 69-76. T. M. Newcomb, E. L. Hartley, and others, Readings in Social Psychology. New York: Henry Holt and Company, 1947. 688 pp.

Burtt, H. E., Principles of Employment Psychology, Revised Edition. New York: Harper & Brothers, 1942. 568 pp.

Cleeton, G. U. and C. W. Mason, Executive Ability, Its Discovery and Development. Yellow Springs, Ohio: Antioch Press, 1946. 540 pp.

Cronbach, L. J., Essentials of Psychological Testing. New York: Harper & Brothers, 1949. 475 pp.

Ghiselli, E. E. and C. W. Brown, Personnel and Industrial Psychology, First Edition. New York: McGraw-Hill Book Company, Inc., 1948. 475 pp.

Greene, E. B., Measurements of Human Behavior. New York: Odyssey Press, 1941. 777 pp.

Guilford, J. P., Psychometric Methods, First Edition. New York: McGraw-Hill Book Company, Inc., 1936. 566 pp.

Halsey, G. D., Making and Using Industrial Service Ratings, First Edition. New York: Harper & Brothers, 1944. 149 pp.

Hollingworth, H. L., Judging Human Character. New York: D. Appleton and Company, 1928. 268 pp.

Jucius, M. J., Personnel Management. Chicago: Richard D. Irwin, Inc., 1947. 696 pp.

Krech, D. and R. S. Crutchfield, Theory and Problems of Social Psychology, First Edition. New York: McGraw-Hill Book Company, Inc., 1948. 639 pp.

Lawshe, C. H., Principles of Personnel Testing, First Edition. New York: McGraw-Hill Book Company, Inc., 1948. 227 pp.

Levine, J. M. and G. Murphy, "The Learning and Forgetting of Controversial Material", pp. 103-113. T. M. Newcomb, E. L. Hartley, and others, Readings in Social Psychology. New York: Henry Holt and Company, 1947. 688 pp.

Adkins, [illegible]
men [illegible]
[illegible]

Bartlett, [illegible]
Hancock, [illegible]
Psychology. [illegible]

Burtt, [illegible]
Edition.

Claxton, [illegible]
and [illegible]
[illegible]

Cronbach, [illegible]
Harper [illegible]

Dallenbach, [illegible]
circle [illegible]
Penn, [illegible]

Freeman, [illegible]
Odyssey [illegible]

Guilford, [illegible]
York: [illegible]

Melaney, [illegible]
Tiers [illegible]

Rolling, [illegible]
Appleton.

Tuolma, [illegible]
Inc., [illegible]

Troop, [illegible]
Psychology [illegible]
Penn, [illegible]

Lawlis, [illegible]
Mc [illegible]

Levine, [illegible]
[illegible]

BIBLIOGRAPHY

Mosher, W. E. and J. D. Kingsley, Public Personnel Administration, Revised Edition. New York: Harper & Brothers, 1941. 671 pp.

Naval Leadership, First Edition. Annapolis, Maryland: U. S. Naval Institute, 1943. 324 pp.

Pigors, P. and C. A. Myers, Personnel Administration a Point of View and a Method, First Edition. New York: McGraw-Hill Book Company, Inc., 1947. 553 pp.

Probst, J. B., Service Ratings. (Baltimore: Lord Baltimore Press, 1951. 94 pp.

Probst, J. B., Measuring and Rating Employee Value. New York: The Ronald Press Company, 1947. 166 pp.

Scott, W. D., R. C. Clothier, S. B. Mathewson, and W. R. Spriegel, Personnel Management, Third Edition. New York: McGraw-Hill Book Company, Inc., 1941. 589 pp.

Stagner, R., Psychology of Personality, Second Edition. New York: McGraw-Hill Book Company, Inc., 1948. 485 pp.

Thorndike, E. L., Individuality. Boston: Houghton Mifflin Company, 1911. 56 pp.

Tiffin, J., Industrial Psychology, Second Edition. New York: Prentice-Hall, Inc., 1947. 553 pp.

White, L. D., Introduction to the Study of Public Administration, Third Edition. New York: The Macmillan Company, 1948. 612 pp.

Yoder, D., Personnel Management and Industrial Relations. New York: Prentice-Hall, Inc., 1948. 845 pp.

B. PERIODICAL ARTICLES

Armstrong, T. O., "Talking Your Ratings", Personnel, 20: 112-115, 1943.

Bingham, W. V., "Halo Valid and Invalid", Journal of Applied Psychology, 23: 221-228, 1939.

Bittner, R. H., "Developing an Industrial Merit Rating Procedure", Personnel Psychology, 1: 403-432, 1948.

BIBLIOGRAPHY

Moeller, [illegible] and [illegible]. [illegible], [illegible] [illegible]-
tration, Revised Edition. New York: [illegible],
1941. 171 pp.

Naval Leadership, First Edition. [illegible], [illegible], [illegible].
Naval Institute, 1943. 320 pp.

Rigors, P. and [illegible]. [illegible], *Personnel Administration: A Point*
of View and a Method. First Edition. New York: McGraw-
Hill Book Co., Inc., 1947. [illegible] pp.

Probst, G. [illegible], *Service Training*. (Mimeographed Lecture
Press, 1951. [illegible] pp.

Probst, G. [illegible], *[illegible]*. [illegible]
The [illegible] Press [illegible] pp.

Scott, [illegible], [illegible], [illegible]
Psychological Problems in Industry, [illegible] Edition.
Modern-Hill [illegible] pp.

Stagner, R., *Psychology of Personality*. [illegible] Edition. New
York: McGraw-Hill [illegible] 1948.

Thorndike, [illegible], *Educational Psychology*. [illegible]
Columbia, 1913. 86 pp.

Tiffin, J., *Industrial Psychology*, [illegible] Edition.
Prentice-Hall, Inc., 1947. 501 pp.

White, L., *Types of [illegible] Among [illegible]*
[illegible]. [illegible] Edition. New York: The Macmillan Company,
1948. 211 pp.

Yoder, D., *Personnel Management and Industrial Relations*. [illegible]
York: [illegible], [illegible], 1948. [illegible] pp.

[illegible]

Armstrong, [illegible], "[illegible]", *[illegible]*, [illegible]
[illegible], 1944.

Bingham, [illegible], [illegible], *[illegible]*
[illegible]. [illegible], 1944.

Gilbreth, [illegible], "[illegible]", *[illegible]*
[illegible], *[illegible]*, [illegible], 1[illegible]-[illegible].

BIBLIOGRAPHY

Bittner, R. H., "Developing an Employee Merit Rating Procedure", Personnel, 25: 278-290, 1949.

Cagle, M. W., Lieutenant Commander, USN., "The Navy's Fitness Report System", United States Naval Institute Proceedings, 73: 779-781, 1947.

Driver, R. S., "Training as a Means of Improving Employee Performance Rating", Personnel, 18: 364-370, 1942.

Ewart, E., S. L. Seashore, and J. Tiffin, "A Factor Analysis of an Industrial Merit Rating Scale", Journal of Applied Psychology, 25: 481-486, 1941.

Flanagan, J. C., "A New Approach to Evaluating Personnel", Personnel, 26: 35-42, 1949.

Hay, E. N., "Selling a Rating Plan", Personnel, 18: 42-48, 1941.

Heider, F., "Social Perception and Phenomenal Causality", Psychological Review, 51: 358-374, 1944.

Hill, R. L., "Efficiency Ratings", Personnel Journal, 15: 330-332, 1937.

Hurlock, E. B., "An Evaluation of Certain Incentives Used in School Work", Journal of Educational Psychology, 16: 145-159, 1925.

Knauft, E. B., "Construction and Use of Weighted Check List Rating Scales for Two Industrial Situations", Journal of Applied Psychology, 32: 63-70, 1948.

Knowles, A. S., "Merit Rating and Labor Management", Personnel, 17: 29-42, 1940.

Knowles, A. S., "Merit Rating of Supervisors, Foremen, and Department Heads", Personnel, 17: 117-133, 1940.

Lawshe, C. H., N. C. Kephart, and E. J. McCormick, "The Paired Comparison Method for Rating Performance of Industrial Employees", Journal of Applied Psychology, 33: 69-77, 1949.

Leuba C. and C. Lucas, "The Effects of Attitudes on Descriptions of Pictures", Journal of Experimental Psychology, 35: 517-526, 1945.

MacCullough, A. V., "Periodic Review of Management Personnel," Personnel, 24: 140-149, 1947.

BIBLIOGRAPHY

Bittner, [illegible] "Developing an Empirical [illegible] Procedure", Personnel, 26: 272-280, 1948.

Cagle, H. W., Lieutenant Commander, [illegible] "The Fitness Report System", United States Naval Institute Proceedings, 73: 779-781, 1947.

Driver, R. C., "Training a Corps of Supervisors: Supervisor Performance Rating", Personnel, 18: 364-370, 1942.

Ewart, E. S., [illegible] and A. Griffin, "A Study [illegible] of an Industrial Training Unit", Journal of Applied Psychology, 26: 481-494, 1942.

Flanagan, J. C., "A New Approach to Validating Forced Choice", Personnel, 26: 35-44, 1946.

Hay, R. H., "Selling a Rating Plan", Personnel, 18: [illegible], 1942.

Helder, F., "Social Perception and Psychological Ability", Psychological Review, 21: 323-334, 1944.

Hill, R. L., "Efficiency Ratings", Personnel Journal, 16: 330-338, 1937.

Murdock, H. F., "An Analysis of Current Selective Service School Data", Journal of Educational Psychology, 34: 142-150, 1943.

Knauft, E. B., "Consensus [illegible] Applied Psychology, 29: [illegible], 1945.

Knowlen, [illegible] "Merit [illegible] Personnel, 17: [illegible], 1940.

Knowles, [illegible] "Merit [illegible] Personnel, [illegible], 1942.

Lawshe, [illegible] Comparison [illegible] Psychol. Monogr., [illegible], 1945.

Lembo, [illegible] "On Measuring Morale", [illegible] Sociol., [illegible], 1944.

MacDiarmid, [illegible] "Employee Merit [illegible] Personnel, [illegible], 1942.

BIBLIOGRAPHY

Marble, S. D., "A Performance Basis for Employee Evaluation", Personnel, 18: 217-226, 1942.

Murray, H. A., "The Effect of Fear Upon Estimates of Maliciousness of Other Personalities", Journal of Social Psychology, 4: 310-329, 1933.

Richardson, M. W. and G. F. Kuder, "Making a Rating Scale That Measures", Personnel Journal, 12: 36-40, 1933.

Rundquist, E. A. and M. H. Bittner, "A Merit Rating Procedure Developed by and For the Raters", Personnel, 26: 273-283, 1950.

Shartle, C. L., "Leadership and Executive Performance", Personnel, 25: 370-380, 1949.

Sisson, D. E., "Forced Choice--The New Army Rating System", Personnel Psychology, 1: 365-381, 1948.

Stockford, L. and H. V. Bissell, "Factors Involved in Establishing a Merit Rating Scale", Personnel, 26: 85-110, 1949.

Symonds, P. M., "On the Loss of Reliability in Ratings Due to Coarseness of the Scale", Journal of Experimental Psychology, 7: 456-461, 1924.

Thurstone, L. L., "The Method of Paired Comparisons for Social Values", Journal of Abnormal and Social Psychology, 21: 384-400, 1927.

Tiffin, J. and E. Musser, "Weighting Merit Rating Items," Journal of Applied Psychology, 26: 575-583, 1942.

Wherry, R. J. and W. H. Fryer, "Buddy Ratings: Popularity Contest or Leadership Criteria", Personnel Psychology, 2: 147-159, 1949.

C. OFFICIAL PUBLICATIONS

An Evaluation of the Report on the Fitness of Officers, P-4114-dmc, Project No. 427. Washington: Training Standards and Curriculum Division, Bureau of Naval Personnel, Department of the Navy, July, 1945. 13 pp.

Methodological Investigation of the Forced Choice Technique Utilizing the Officer Description and the Officer Evaluation Blanks, Study #701, PR 4061-04. Washington:

BIBLIOGRAPHY

Marble, [illegible]. "A [illegible] Basis [illegible]", Personnel, 10: 217-22[illegible], 1948.

Murray, H. A., "The [illegible] of Personality as a [illegible]-dure of Other Personalities", Journal of Social Psychology, 4: 310-329, 1933.

Richardson, M. [illegible], "On Long a [illegible] Of the That Measures", Educational Journal, [illegible]: 30-[illegible], [illegible].

Rundquist, E. [illegible], "A [illegible] Psychologed [illegible], Psychometrika, [illegible]: [illegible]-[illegible], 1950.

Shartle, C. L., "Leadership [illegible] nal, [illegible]: [illegible], [illegible].

Stason, G. [illegible], "Personal [illegible] Personnel [illegible], [illegible].

Stockford, L. and E. [illegible], "Theory [illegible] ship [illegible], [illegible], [illegible], [illegible].

Symonds, P. M., "On [illegible] Conference of [illegible] Journal of [illegible] chol., 7: 400-[illegible], [illegible].

Thurstone, L. L., "[illegible] Values", Journal of Abnormal and Social Psychology, [illegible]: 3[illegible]-[illegible], [illegible].

Tiffin, J. and [illegible], "[illegible] nal of Applied Psychology, [illegible]: [illegible], [illegible].

Zberry, [illegible] tomic [illegible] Psychology, [illegible], [illegible]: [illegible], [illegible].

BIBLIOGRAPHY

Adjutant General's Office, Department of the Army, 6 July, 1948. 17 pp.

Register of Commissioned and Warrant Officers of the United States Navy and Marine Corps, January 1, 1949. Washington: Government Printing Office. 1447 pp.

The 1948 Standardization of Officer Efficiency Report, WD AGO FORM 67-1, WD AGO PRS Report 799, Washington: Adjutant General's Office, Department of the Army, 7 June, 1949. 28 pp.

United States Navy Regulations, 1948. Washington: Government Printing Office. 403 pp.

D. PARTS OF A SERIES

Allport, G. W. and H. S. Odbert, "Trait Names: A Psycholexical Study", Psychological Monographs, Psychological Review Publications, Volume 47, Number 211. Princeton: Psychological Review Company, 1936. 171 pp.

Bills, M. L., "Merit Rating and Seniority", American Management Association Personnel Series, Number 127. New York: American Management Association, 1945. pp 33-39.

Bradshaw, F. F., "The American Council on Education Rating Scale, Its Reliability, Validity, and Use. New York: Archives of Psychology, Number 119, 1930. 80 pp.

Kidder, J. T., Employee Rating, Studies in Personnel Policy, National Industrial Conference Board, Number 39, 1942. 119 pp.

Tiffin, J., "Merit Rating: Its Validity and Techniques", American Management Association Personnel Series, Number 100. New York: American Management Association, 1945. pp. 14-23.

E. PUBLICATIONS OF LEARNED ORGANIZATIONS

Construction of Three Measures for Instructor Evaluation, Interim Report on Educational Research Project, Report 383-1-5. New York: Richardson, Bellows, Henry, & Company, Inc., 20 July, 1948. 44 pp.

BIBLIOGRAPHY

Adjutant General's Office, Department of the Army, 6 July, 1956. IV pp.

Register of Commissioned and Warrant Officers of the United States Navy and Marine Corps, August 1, 1958. Washington: Government Printing Office. [illegible] pp.

The 1948 Standardization of Officer Efficiency Report, NAVPERS 2600 (Rev. 10-48). [illegible] Washington: Chief of Naval Operations, Navy Department, 1948. 28 pp.

United States Navy Regulations, 1948. Washington: Government Printing Office. [illegible] pp.

B. Part [illegible]

Allport, G. W. and H. S. Odbert, "Trait-names: a psycholexical study", Psychological Monographs, Psychological Review Publications, Volume 47, Number 211. Princeton: Psychological Review-Logical Review Company, 1936. [illegible] pp.

Dills, R. L., "Merit rating as a analysis", Personnel and Guidance Journal, Volume [illegible], Number [illegible], [illegible] 1957. New York: American Personnel and Guidance Association, [illegible]. pp. 23-26.

Strabley, F. T., The Positive Results of Education Series Design, Air Materiel Command, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, [illegible], 1950. 30 p.

Kidder, A. T., Employee Ratings, [illegible] in Personnel Administration, National Industrial [illegible], Chicago, 1948. 110 pp.

Tiffin, J., "Merit rating: [illegible] Industrial Psychology, Prentice-Hall, Inc. New York: [illegible] pp. 1-20.

C. [illegible]

Construction of Forms Designed for Evaluation of Officers, Interim Report on Industrial Psychology Project [illegible], Washington, D. C., [illegible], 1950. [illegible] pp.

BIBLIOGRAPHY

Preston, H. O., The Development of a Procedure for Evaluating Officers in the United States Air Force, Technical Appendices and Notes. Pittsburgh: American Institute For Research, 7 July, 1948. 231 pp.

Personnel Research Board, The Ohio State University, Studies in Naval Leadership, Technical Report. Columbus, Ohio: The Ohio State University Research Foundation, June 15, 1949. 87 pp.

F. UNPUBLISHED MATERIALS

Baze, G. S., A Rating Scale for Marine Non-Commissioned Officers of the First Three Pay Grades, (unpublished Master's thesis, The Ohio State University, Columbus, Ohio, 1947), 76 pp.

Samples, W. G., A Study of Merit Rating Methods for Professional Employees, (unpublished Master's thesis, The Ohio State University, Columbus, Ohio, 1948), 97 pp.

BIBLIOGRAPHY

Preston, H. G., The [illegible] of a [illegible] [illegible] officers in the United States [illegible] Press, [illegible] Press, [illegible] Jesse Van Horne. [illegible] Season, V [illegible], 1948. [illegible] pp.

Personnel Reserve Board, The Ohio Class [illegible], [illegible] in Naval Ice [illegible], [illegible] The Ohio [illegible] Naval Training, [illegible] 1, 1940. 27 pp.

[illegible]

Rano, R. [illegible] of the Naval [illegible] (Ohio [illegible]) The Ohio [illegible], [illegible] pp.

Sampion, [illegible] tional Surplus, (Ohio [illegible] State University, [illegible] State University, Columbus, Ohio, 1946. 27 pp.

APPENDIX A

SAMPLES OF TYPES OF RATING FORMS

APPENDIX A.

Evaluative Type Graphic Scale

Traits				
Intelligence	Excellent	Very Good	Good	Unsatisfactory
Initiative	Unsatisfactory	Good	Very Good	Excellent
Leadership	Excellent	Very Good	Good	Unsatisfactory

A. [illegible]

Evaluative Type Rating Scale

Traits				
Intelligence	Very good	Good	Average	Unsatisfactory
Initiative	Very good	Good	Average	Unsatisfactory
Industrious	Very good	Good	Average	Unsatisfactory

Descriptive Type Rating Scale

Qualities				
Intelligence (With reference to the faculty of comprehension; mental acuteness)	Exceptionally quickwitted; keen in understanding.	Grasps essentials of a situation quickly.	Readily understands normal situations and conditions.	Slow of comprehension; unimaginative.
Initiative (With reference to constructive thinking and resourcefulness)	Requires constant guidance and supervision in his work or evades responsibilities.	Capable of performing normal and routine duties on own responsibility.	Able to plan and execute missions on his own responsibility.	Exceptional in ability to think, plan, and do things without waiting to be told and instructed.
Leadership (With reference to the faculty of directing, controlling, and influencing others in definite lines of action.)	Inspires others to a high degree by precept and example; will always have a following.	A very good leader. Attracts subordinates.	Leads well but has his limitations.	A poor leader of men; fails to attract others; does not inspire confidence.

From the Navy rating form shown in Appendix C

Descriptive Type Rating Scale

APPENDIX A

FORCED DISTRIBUTION SCALE

Traits	Highest 10%	Next 20%	Middle 40%	Next 20%	Lowest 10%
Intelligence					
Initiative					
Leadership					

CHECK LIST SCALE

Instructions to rater: Check only those statements among the following which are applicable to the ratee.

:____: Is quick to take the correct action when it is needed.

:____: Is resentful of constructive criticism.

:____: Can be trusted to carry out assignments to best of his ability.

:____: Is frequently late to work.

:____: Requires constant supervision.

Traits	Highest 10%	Next 20%	Middle 40%	Next 20%	Lowest 10%
Intelligence					
Initiative					
Leadership					

APPENDIX B

AN EARLY NAVY RATING FORM

(Report on Officers under U. S. Navy Regulation Circular No. 86, September 10, 1891. A separate blank to be used for each Officer.)

REPORT ON THE FITNESS OF OFFICERS.

(FORM A.)

Name. . *Rank.*

U. S. S. Commanding U. S. S.

Station of duty. .....

Period covered by this report, from . ..... , 18 ,

to , 18 .

1. Ability to command. .....
2. Manner of performing duties.
3. General conduct. .....
4. Sobriety. .....
5. Health. ...
6. Condition and efficiency of command.
7. If any special duty has devolved upon him, state its nature, and how it was performed.

8. Remarks. .....

.....

.....

.....

.....

.....

.....

.....

.....

I certify on honor that I believe the entries herein made are true and impartial, and that I have communicated unfavorable mention, as directed.

..... , *U. S. N.*,

Comdg.

NOTE.—Under "Remarks," refer to any special reports made in accordance with Par. V, Regulation Circular No. 86, within the period covered by this report.

Answers to queries from 1 to 6 inclusive shall be either "Excellent," "Good," "Tolerable," or "Not good," and must be written by the officer signing the report. Should the answer to any query be "Not good," or of an unfavorable nature, the reasons for such answer must be clearly stated, and a copy of that portion of the report furnished to the officer concerned, who shall be granted a reasonable time to prepare such written statement in regard to it as he may desire to make, which statement shall be endorsed and forwarded with the report. Should the officer not desire to make a statement that fact shall be stated.

APPENDIX C

A MORE RECENT NAVY RATING FORM

12. To what degree has he exhibited the following qualifications? (Each line is divided into four main divisions, with a definition below each division. Consider only the definition below each line which accurately fits in regard to the subject. Having decided which definition is most suitable, place a check mark at the grading along line to which the person being marked is entitled. A check to the right of the double vertical line will be considered as an unsatisfactory (unfavorable) report.)

Intelligence (With reference to the faculty of comprehension; mental acuteness.)	Exceptionally quick-witted; keen in understanding.	Grasps essentials of a situation quickly.	Readily understands normal situations and conditions.	Slow of comprehension; unimaginative
Judgment (With reference to a discriminating perception by which the values and relations of things is mentally asserted.)	Unusually keen in estimating situations and reaching sound decisions.	Can generally be depended on to make proper decisions.	Good judgment in normal and routine things.	Poor judgment; draws wrong conclusions.
Initiative (With reference to constructive thinking and resourcefulness; ability and intelligence to act on own responsibility.)	Exceptional in ability to think, plan, and do things without waiting to be told and instructed.	Able to plan and execute missions on his own responsibility.	Capable of performing normal and routine duties on own responsibility.	Requires constant guidance and supervision in his work, or evades responsibilities.
Force (With reference to moral power possessed and exerted in producing results.)	Strong, dynamic.	Strong.	Effectual under normal and routine circumstances.	Weak; with little influence over others.
Leadership (With reference to the faculty of directing, controlling, and influencing others in definite lines of action.)	Inspires others to a high degree by precept and example; will always have a following.	A very good leader; attracts subordinates.	Leads well, but has his limitations.	A poor leader of men; fails to attract others; does not inspire confidence.
Tact (With reference to the faculty of being courteous, considerate, and sensible in dealing with others.)	Adaptable and courteous to a high degree. Can be relied upon to handle delicate questions or situations.	Adaptable, courteous, and successful in dealing with others.	Reasonably tactful.	Irritates; creates friction.
Cooperation (With reference to the faculty of working harmoniously with others toward the accomplishment of common duties.)	Exceptionally successful in working with others to a common end.	Works in harmony with others.	Responsive to leadership.	Obstructive.
Loyalty (Fidelity, faithfulness, allegiance, constancy — all with reference to a cause and to higher authority.)	Unswerving in allegiance; frank and honest in aiding and advising.	A high sense of loyalty.	Faithful in the execution of his duty.	Hypercritical; disdainful.
Attention to Duty (With reference to the faculty of carrying on all work in a thorough and conscientious manner.)	Has an exceptionally high standard with reference to devotion to duty.	Attends to his duty in a highly satisfactory manner.	Satisfactory attention to duty.	Inclined to defer or neglect his work.
Presence of Mind (With reference to the faculty of acting instinctively in a logical manner in difficult and unforseen situations.)	Exceptionally cool-headed and logical in his actions under all conditions.	Composed and logical in his actions in difficult situations.	Logical in his actions in general.	Easily disconcerted.
Endurance (With reference to ability for carrying on under any and all conditions.)	Capable of standing an exceptional amount of physical hardships and strain.	Can perform well his duties under trying conditions.	Of normal endurance.	Tires or breaks down easily.
Industry (With reference to performance of duties in an energetic manner.)	Extremely energetic and industrious.	Thorough and energetic.	Intermittently energetic and industrious.	Indolent; inactive; lazy.
Military Bearing (With reference to military carriage and dignity of demeanor.)	Exceptional in carriage.	Carries himself well.	Bearing fair.	Unmilitary in bearing.
Neatness of Person and Dress (With reference to correctness of uniform and smartness of appearance.)	Punctilious as to uniform; immaculate in dress and person.	Properly and well dressed and careful of personal appearance.	Generally neat and correct in uniform.	Untidy in dress or personal appearance.
Aptitude for Service (With reference to a natural capacity and general fitness for the Naval Service.)	Enthusiastic and excellently fitted for the Naval Service.	Well fitted and interested.	Adapts himself to and is acquiring a capacity for the service.	Not fitted.

REMARKS

13. Give in this space: (*a*) A clear and concise estimate of this officer's personal and military character, (*b*) his fitness for promotion, (*c*) any duty performed worthy of special mention, (*d*) also any information which might be of value to the department in making assignments to duty.

THIS SPACE MUST NOT BE LEFT BLANK
(See Art. 137 (7), U. S. N. R., 1920)

APPENDIX D

PRESENT NAVY RATING FORM

OFFICER'S FITNESS REPORT WORK SHEET

WORK SHEET

READ CAREFULLY THE INSTRUCTIONS ON THE REVERSE SIDE

DATE

1. NAME (last) (first) (middle) | RANK AND CLASSIFICATION | FILE NO.

SHIP OR STATION | PERIOD OF REPORT (mo., day, year) DATE FROM | DATE TO

DATE OF REPORTING TO PRESENT SHIP OR STATION | OCCASION FOR REPORT ☐ DETACHMENT OF OFFICER REPORTED ON ☐ DETACHMENT OF REPORTING SENIOR ☐ REGULAR SEMI-ANNUAL ☐ QUARTERLY ☐ SPECIAL

2. DESCRIPTION OF DUTIES SINCE LAST FITNESS REPORT (List most recent first and describe accurately. Include periods of leave, transit, etc., also include employment of ship.)

	FROM MO.	FROM YR.	TO MO.	TO YR.

Has present duty changed since last fitness report was submitted? ☐ Yes ☐ No

3. IF COURSES OF INSTRUCTION WERE COMPLETED DURING PERIOD OF THIS REPORT, LIST TITLE OF COURSE, LOCATION OF SCHOOL, LENGTH OF COURSE AND DATE COMPLETED.

Are you physically qualified for Sea Duty? ☐ Yes ☐ No ☐ Don't Know

4. If Aviator, indicate No. of Flight Hours Last Two years for Each Type Aircraft (List Most Recent Type First)

TYPE OF AIRCRAFT							TOTAL
NO. OF HOURS							

5. MY PREFERENCE FOR NEXT DUTY IS:

SEA	KIND OF DUTY	LOCATION
SHORE	KIND OF DUTY	LOCATION

6. *SECTIONS 6 THROUGH 12 TO BE FILLED IN BY REPORTING OFFICER* | NAME OF REPORTING OFFICER | RANK | FILE NO. | OFFICIAL STATUS RELATIVE TO OFFICER REPORTED ON

IS THIS OFFICER QUALIFIED TO PERFORM ALL HIS PRESENT DUTIES? ☐ YES ☐ NO | INDICATE MORE RESPONSIBLE DUTIES FOR WHICH HE IS IN TRAINING. (If none, so state) | DATE OF EXPECTED QUALIFICATION

APPRAISAL FOR DETAIL PURPOSES

Comment on special or outstanding qualifications as well as any physical defects, which should be considered in determining the kinds of duty to which he should be detailed. Only comments on qualifications significant in detailing should be entered here. ANY COMMENTS REGARDING FITNESS FOR PROMOTION SHOULD BE ENTERED IN SECTION 12 ONLY OF PAGE 1.

FOR WHAT DUTIES IS HE RECOMMENDED?

ASHORE AFLOAT

7. FOR EACH FACTOR OBSERVED CHECK THE APPROPRIATE BOX TO INDICATE HOW THE OFFICER COMPARES WITH ALL OTHERS OF THE SAME RANK, CLASSIFICATION AND CORPS WHOSE PROFESSIONAL ABILITIES ARE KNOWN TO YOU PERSONALLY. DO NOT LIMIT THIS COMPARISON ONLY TO THE OTHERS NOW UNDER YOUR COMMAND. DO NOT HESITATE TO MARK "NOT OBSERVED" ON ANY QUALITY WHEN APPROPRIATE. NO ENTRY WHICH IS MADE IN THIS SECTION WILL BE CONSIDERED AN UNSATISFACTORY REPORT WHICH MUST BE REFERRED TO THE OFFICER FOR STATEMENT. ONLY ENTRIES DESIGNATED IN SECTIONS 8, 9, 11 AND 12 WILL BE SO CONSIDERED.

RATING FACTORS		Not Observed	Within Bottom 10%	Within Next 20%	Within Middle 40%	Within Next Top 20%	Within Top 10%
A. SEA OR ADVANCE BASE DUTY How does this officer compare in: NOTE: ITEM (A3) TO BE MARKED FOR ALL OFFICERS.	1. STANDING DECK WATCHES UNDERWAY?						
	2. ABILITY TO COMMAND?						
	3. PERFORMANCE IN PRESENT DUTIES AS DESCRIBED IN SECTION 2, ABOVE?						
	4. REACTIONS DURING EMERGENCIES?						
	5. PERFORMANCE AT BATTLE STATION OR IN BATTLE DUTIES?						
B. INITIATIVE AND RESPONSIBILITY How well does this officer:	1. ASSUME RESPONSIBILITY WHEN SPECIFIC INSTRUCTIONS ARE LACKING?						
	2. GIVE FRANK OPINIONS WHEN ASKED OR VOLUNTEER THEM WHEN NECESSARY TO AVOID MISTAKES?						
	3. FOLLOW THROUGH DESPITE OBSTACLES IN CARRYING OUT RESPONSIBILITIES ASSIGNED OR ASSUMED?						
C. UNDERSTANDING AND SKILL How well does this officer:	1. UNDERSTAND INSTRUCTIONS GIVEN, AND USE SUGGESTIONS OFFERED?						
	2. EXERCISE JUDGMENT?						
	3. RATE IN TECHNICAL COMPETENCE IN HIS SPECIALTY, IF ANY? (Name Specialty)						
D. LEADERSHIP How well does this officer:	1. INSPIRE SUBORDINATES TO WORK TO THE MAXIMUM OF THEIR CAPACITY?						
	2. EFFECTIVELY DELEGATE TASKS AND AUTHORITY?						
	3. TRANSMIT ORDERS, INSTRUCTIONS, AND PLANS?						
	4. ORGANIZE HIS WORK AND THAT OF THOSE UNDER HIS COMMAND OR SUPERVISION?						
	5. MAINTAIN DISCIPLINE AMONG THOSE UNDER HIS COMMAND OR DIRECTION?						
E. CONDUCT AND WORK HABITS How does this officer compare in:	1. ABILITY TO WORK WITH OTHERS?						
	2. ABILITY TO ADAPT TO CHANGING NEEDS AND CONDITIONS?						
	3. MILITARY CONDUCT—BEARING, DRESS, COURTESY, ETC.?						

8. INDICATE YOUR ATTITUDE TOWARD HAVING THIS OFFICER UNDER YOUR COMMAND, WOULD YOU: (Check one) ☐ DEFINITELY NOT WANT HIM? (UNSATISFACTORY) ☐ PREFER NOT TO HAVE HIM? (UNSATISFACTORY) ☐ BE SATISFIED TO HAVE HIM? ☐ BE PLEASED TO HAVE HIM? ☐ PARTICULARLY DESIRE HIM?

9a. Considering All Officers of the Same Rank, Classification and Corps, Whose Professional Abilities Are Known to You Personally, Would You Promote Him: (Check one) ☐ UNDER NO CIRCUMSTANCES? (Unsatisfactory) ☐ IF 90% WERE TO BE PROMOTED? ☐ IF 70% WERE TO BE PROMOTED? ☐ IF 30% WERE TO BE PROMOTED? ☐ IF ONLY 10% WERE TO BE PROMOTED?

9b. How many Officers are included in the group used for the comparison in 9a? ☐ 10 OR LESS ☐ 10 TO 50 ☐ OVER 50

10. COMMENT IN SECTION 12 AND GIVE REFERENCE HERE TO ANY COMMENDABLE OR ADVERSE REPORTS THAT HAVE BEEN MADE ON THE OFFICER DURING THIS PERIOD.

11. HAVE YOU ANY ADVERSE COMMENTS TO MAKE REGARDING THIS OFFICER'S QUALITIES OR PERFORMANCE? ☐ YES ☐ NO
HAS HE ANY MENTAL OR MORAL WEAKNESS WHICH ADVERSELY AFFECTS HIS EFFICIENCY? ☐ YES ☐ NO
If yes, explain in Section 12.

UNSATISFACTORY. Yes in either item of Section 11 constitutes an unsatisfactory report and must be referred to the officer for statement.

12. Give in this space a clear, concise appraisal of the officer reported on and his performance of duty, including any worthy of special mention. Include recommendations as to promotion. Any statements of unsatisfactory performance, ability, character, or conduct must be referred to the officer for statement. Statements of a constructive nature which refer to minor imperfections or lack of qualifications do not constitute an unsatisfactory report: For example: "This officer was a little slow in getting started but is now making good progress" or "This officer is well qualified in his present duties but has had no experience at sea" would not be unsatisfactory in nature.

Check one of these boxes — I CONSIDER THIS REPORT TO BE ☐ SATISFACTORY ☐ UNFAVORABLE ☐ UNSATISFACTORY

DO NOT LEAVE BLANK

(If additional space is needed attach extra sheet)

SIGNATURE OF OFFICER REPORTED ON (Applies only to Sections 1 through 8) | SIGNATURE OF REPORTING OFFICER | Has this report been shown or referred to officer reported on? ☐ Yes ☐ No

DO NOT FORWARD THIS SHEET TO BUPERS

152

OFFICER'S FITNESS REPORT
INSTRUCTION SHEET

IMPORTANT INSTRUCTIONS—READ CAREFULLY

GENERAL INSTRUCTIONS

The attached revised Officer's Fitness Report is to be used in place of the old forms, NAVPERS 310 and 311.

This form serves the following purposes:

1. It serves as a report of fitness for all officers both afloat and on shore.
2. The first carbon—(Page 2)—keeps up to date in BuPers the Officer's Qualifications Questionnaire, which provides the Bureau with information covering each officer's previous experience and qualifications for various types of duty.
3. The second carbon—(Page 3)—provides data covering changes in the officer's qualifications and is to be filed in the Officer's Qualification Record Jacket as an aid to Commanding Officers and Personnel Officers in assigning him properly.

This form is to be submitted semi-annually for all officers and in all cases of permanent detachment of either the officer or reporting senior. Special reports of fitness on an officer, on the prescribed form, shall be made whenever the officer reported on:

(a) Distinguishes himself in battle.

(b) Performs an outstanding act of valor or devotion to duty.

(c) Displays extraordinary courage, ability, or resource in time of peril or great responsibility.

(d) Is guilty of serious misconduct or marked inefficiency.

A typewriter is to be used when at all possible in filling out Sections 1 through 6. Since 96% of all fitness reports received in BuPers are typed, the form has been constructed for that type of preparation. Care should be exercised that the carbon copies are legible if a typewriter is not used.

INSTRUCTIONS FOR REPORTING OFFICERS

In deciding on promotions of officers, Selection Boards must, in effect, compare an officer with others of the same rank rather than with more arbitrary standards. You will note that in Section 7 and subsequent sections you are asked to do just that — compare each officer with all others of the same rank and corps whose professional abilities are known to you personally. Please note that the officer is not to be compared only with the others of his rank now under your command. For this reason, it is important to indicate in Section 9b how many officers are included in the group you use for comparison.

In making this comparison, keep in mind that the group of officers whose professional abilities are known to you personally (or any other group of people) will fall into a normal distribution when graded on any trait or factor—that is, there will be a small number at the lower end, a larger group in the middle, and a small group at the top. With this curve in mind, compare the officer with the group and mark him on each factor in Section 7 as falling in one of the five brackets—the lower 10%, the next 20%, the middle 40%, the next 20% or the top 10%. Do not hesitate to mark "not observed" on any factor which you think not applicable to the duty in which you have observed the officer or in which your observation has been too limited to warrant judgment.

No entry which is made in Section 7 will be considered an unsatisfactory report. Only adverse comment in Section 6 and entries so designated in Sections 8, 9, 11 and 12 will be so considered.

An unsatisfactory report must be referred to the officer reported on for his statement which is to be attached to the report of fitness. In any case open to question as to what constitutes an entry of an unfavorable or unsatisfactory nature the officer will always be given the benefit of having seen the report. (See Articles 1701 and 1405 Navy Regulations, and BuPers Manual Article B-2202.

The Bureau desires that reporting seniors make every effort to show each fitness report to the officer reported upon and to discuss it with him, in so far as practicable. In this connection please note the instructions in Section 12 which provide that statements of a constructive nature which refer to minor imperfections or lack of qualifications do not constitute an unsatisfactory report. On every report of fitness, the reporting senior will indicate under Section 12 whether the officer reported on has or has not seen the report.

The reporting senior will sign all three pages of the report in the lower right hand corner, or will sign the original and designate a commissioned officer, preferably senior to the officer reported on, to authenticate Pages 2 and 3 in lower right hand corner. The officer reported on may sign and retain Page 3, inserting same in his qualification jacket, if he is geographically detached from the reporting senior.

The Officer's Fitness Report (Page 1) and the Officer's Qualification Report—BuPers Copy—(Page 2) are to be forwarded—not separated—to BuPers. The Officer's Qualification Report—Jacket Copy—(Page 3) is to be detached—and filed in the Officer's Qualification Record Jacket.

Fitness Reports are to be submitted promptly and their preparation is one of the most important and responsible duties of superior officers. Failure to prepare them objectively is detrimental to the efficiency of the Navy. If not submitted promptly, the rights of the officer reported on may be prejudiced. The fitness of an officer for the service with respect to promotion and assignment to duty is determined by his record.

INSTRUCTIONS FOR OFFICER REPORTED ON

It is your responsibility to fill out Sections 1 through 5 of this form and to sign all sheets in the lower left-hand corner. Submit the form to your reporting senior at the times specified in the General Instructions above. Use a typewriter, if at all possible—if not, use ink, but be sure that all copies are legible.

NOTE: For convenience there is printed on the back of these instructions a work sheet which may be used as a draft in preparing the carbonized set. The work sheet is to be detached before filling out the carbonized set and is NOT to be forwarded to BuPers.

APPENDIX E

ARMY RATING FORM

EFFICIENCY REPORT

See AR 600–185 for details.

Unit Adjutant or Personnel Officer will complete Sections I and III.
Rating Officer will complete Sections II, IV, V, VI, VII, VIII, and IX.
Indorsing Officer will complete Sections II, V, VII, and IX.

Section I. OFFICER REPORTED UPON

Use typewriter or print in ink. Use carbon paper to fill out Section III at same time. See AR 600–185.

DO NOT WRITE IN THIS SPACE

LAST NAME	FIRST NAME	INITIAL	SERIAL NUMBER	GRADE	ARM OR SERVICE	COMPONENT	PERIOD OF REPORT FROM	PERIOD OF REPORT TO

THEATER OR CONTINENTAL COMMAND	UNIT, ORGANIZATION, AND STATION	PRIMARY MOS	DUTY ASSIGNMENT (MOS CODE)	DAYS OF DUTY	DAYS OF LEAVE	DAYS OF OTHER NON-DUTY	JP

DATE OF REPORT	FOR REPORTS RENDERED BECAUSE OF PERMANENT CHANGE OF STATION, SUPPLY ADDRESS OF UNIT AND INSTALLATION WHERE OFFICER WILL REPORT	PQ

NAME, GRADE, AND ORGANIZATION OR UNIT OF RATING OFFICER	NAME, GRADE, AND ORGANIZATION OR UNIT OF INDORSING OFFICER	OA

Section II. DATA AND SUGGESTIONS FOR USE IN ASSIGNMENT

NOTE: Information on this page will be forwarded to the Career Branch of the Personnel and Administration Division by TAG after ratings have been determined. Proper future assignment and utilization of the officer will depend upon the care with which information in this section is formulated and reported. Use typewriter or print in ink.

A. DUTIES ACTUALLY PERFORMED ON PRESENT JOB. To be supplied by Rater. Be specific. Give his duty assignment and all additional duties with enough specific detail to show scope of job in each area.

B. DESCRIPTION OF OFFICER RATED AND COMMENTS. These paragraphs should cover physical, mental, moral qualities of rated officer, specialties of value to the Army, and any special defects or weaknesses affecting his ability to do certain assignments.

COMMENTS OF RATING OFFICER	COMMENTS OF INDORSING OFFICER

C. ESTIMATED DESIRABILITY IN VARIOUS CAPACITIES. Assume you are a commander of a major unit in war. Indicate to what extent you would want the rated officer to serve under you in the next higher grade in each type of duty described below. Place an X in the proper box, using the shaded NA area if the duty is not applicable. If line h is used, specify the nature of the specialty.

	RATER NA	RATER 1	RATER 2	RATER 3	RATER 4	RATER 5	INDORSER NA	INDORSER 1	INDORSER 2	INDORSER 3	INDORSER 4	INDORSER 5
a. Represent your viewpoint and make decisions in your name at a higher headquarters.												
b. Command a unit immediately subordinate to you on a combat mission.												
c. Be responsible in an emergency calling for initiative, coolness, forceful leadership.		NOT WANT HIM	TAKE A CHANCE ON HIM	HAPPY TO HAVE HIM	PREFER HIM TO MOST	FIGHT TO GET HIM		NOT WANT HIM	TAKE A CHANCE ON HIM	HAPPY TO HAVE HIM	PREFER HIM TO MOST	FIGHT TO GET HIM
d. Work on an assignment requiring great attention to detail and routine.												
e. Plan all aspects of a military situation, using judgment, initiative, and coolness.												
f. Carry out an assignment in a civilian component such as ROTC, NG, or ORC.												
g. Represent you where tact and ability to get along with people are needed.												
h. Work on an assignment as specialist or technician. (Specify.)												
i. Carry out the duties of the type of work to which he is now assigned.												

D. IMMEDIATE RECOMMENDATIONS FOR CAREER DEVELOPMENT. Be specific.

RATER'S RECOMMENDATION FOR ASSIGNMENT (MOS CODE)	INDORSER'S RECOMMENDATION FOR ASSIGNMENT (MOS CODE)
RATER'S RECOMMENDATION FOR FURTHER TRAINING	INDORSER'S RECOMMENDATION FOR FURTHER TRAINING

E. ENTRIES ARE BASED ON → (RATER WILL CHECK)	INTIMATE DAILY CONTACT	FREQUENT OBSERVATION OF THE RESULTS OF HIS WORK	INFREQUENT OBSERVATION OF THE RESULTS OF HIS WORK	ACADEMIC RECORDS	OFFICIAL REPORTS

F 1429 REV. 3

EFFICIENCY REPORT

WD AGO Form 67-1 Part 2 See AR 600–185 for details.

Unit Adjutant or Personnel Officer will complete Sections I and III.
Rating Officer will complete Sections II, IV, V, VI, VII, VIII, and IX.
Indorsing Officer will complete Sections II, V, VII, and IX.

Section III. OFFICER REPORTED UPON

Enter same information as for Section I.

DO NOT WRITE IN THIS SPACE

LAST NAME	FIRST NAME	INITIAL	SERIAL NUMBER	GRADE	ARM OR SERVICE	COMPONENT	PERIOD OF REPORT FROM	PERIOD OF REPORT TO

THEATER OR CONTINENTAL COMMAND	UNIT, ORGANIZATION, AND STATION	PRIMARY MOS	DUTY ASSIGNMENT (MOS CODE)	DAYS OF DUTY	DAYS OF LEAVE	DAYS OF OTHER NON-DUTY	JP

DATE OF REPORT	FOR REPORTS RENDERED BECAUSE OF PERMANENT CHANGE OF STATION, SUPPLY ADDRESS OF UNIT AND INSTALLATION WHERE OFFICER WILL REPORT	PQ

OA

Section IV. JOB PROFICIENCY

READ INSTRUCTION SHEET CAREFULLY BEFORE MARKING THIS SECTION

	MOST	LEAST
1		
A. Becomes dogmatic about his authority.		
B. Careless & slipshod in attention to duty.		
C. No one ever doubts his ability.		
D. Well-grounded in all phases of Army life.		
2		
A. Follows closely directions of higher echelons.		
B. Inclined to "gold-brick."		
C. Criticizes unnecessarily.		
D. Willing to accept responsibility.		
3		
A. A go-getter who always does a good job.		
B. Cool under all circumstances.		
C. Doesn't listen to suggestions.		
D. Drives instead of leads.		
4		
A. Always criticizes, never praises.		
B. Carries out orders by "passing the buck."		
C. Knows his job and performs it well.		
D. Plays no favorites.		
5		
A. Constantly striving for new knowledge and ideas.		
B. Businesslike.		
C. Apparently not physically fit.		
D. Fails to use good judgment.		
6		
A. Cannot assume responsibility.		
B. Knows how and when to delegate authority.		
C. Offers suggestions		
D. Too easily changes his ideas.		
7		
A. Fails to work for the best interest of all.		
B. Has a high degree of initiative.		
C. Never makes excuses for his mistakes.		
D. Slow in accomplishing his work.		
8		
A. Criticizes policies of superiors.		
B. Others can't work with him.		
C. If he is wrong, will admit it.		
D. The men know they can rely on his judgment.		
9		
A. Doesn't try to "pull rank."		
B. Knows men, their capabilities & limitations.		
C. Low efficiency.		
D. Uses a steady monotone in his speech.		
10		
A. Fails to support fellow officers.		
B. Oversteps his authority.		
C. Gives clear and concise directions.		
D. Very exacting in all details.		
11		
A. Blames others for his mistakes.		
B. Always demands strict discipline.		
C. Excellent at constructive criticism.		
D. Hesitant about rendering decisions.		
12		
A. Can take over in an emergency.		
B. Fair and just in his dealings.		
C. Lacks interest in his job.		
D. Questions orders from superiors.		

Section V. JOB PROFICIENCY

READ INSTRUCTION SHEET CAREFULLY BEFORE MARKING THIS SECTION

DO NOT WRITE IN THIS SPACE

1. Management and operation of military matters not included in tactics and strategy.
2. The direction of the over-all operation of a military unit.
3. Presenting learning materials in a classroom situation in a military or civilian component.
4. Exercise of specialized knowledge, requiring lengthy technological training.
5. Assisting commanders of battalions or larger units in devising methods of meeting the requirements of military situations.
6. Duties involving aeronautical skills performed by rated officers.
7. Training at service schools, Air University, Army Industrial College, etc.

	FOR RATING OFFICER	FOR INDORSING OFFICER
PRIMARY	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10
SECONDARY	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10

DO NOT WRITE IN THIS SPACE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32

1	**4**	**7**	**10**
A. People work for & with him because of his personality.	A. Lacks ability to inspire confidence of men & officers.	A. Plenty of military snap, bearing, & neatness.	A. Obtains respect & obedience without causing resentment.
B. Never rank-conscious.	B. Easygoing.	B. Normally cheerful.	B. Lacks aggressiveness.
C. Thinks only of himself.	C. Type of man everyone likes for a friend.	C. Can't take criticism.	C. Has an excellent command of language.
D. Worries a great deal.	D. Has a quiet, dignified bearing.	D. Doesn't get along with people.	D. Lacking in good conduct & moral habits.
2	**5**	**8**	**11**
A. Active in athletics.	A. Hot-tempered.	A. Modest & reserved.	A. Coolheaded.
B. Firm but not overbearing.	B. Fails to demonstrate originality.	B. Doesn't have drive or force he should.	B. Commands respect by his actions.
C. Egotistical.	C. Reserved.	C. Antisocial.	C. Overbearing.
D. Rubs people the wrong way.	D. Impresses people favorably.	D. Respected by all fellow officers.	D. Indifferent.
3	**6**	**9**	**12**
A. Compliments a man on his good work.	A. Boastful.	A. A quiet, unassuming officer.	A. Immature.
B. Loses his head, gets excited.	B. Inspires pride in the organization.	B. Follows rather than leads.	B. Modest but not retiring.
C. Has admiration of officers & men alike.	C. Lacks tact.	C. Has an attitude of superiority.	C. Nervous.
D. Poor in dress & appearance.	D. Thoughtful of others.	D. Tactful.	D. Thoroughly cooperative in his work.

Section VII. PERSONAL QUALIFICATIONS

Use ELECTROGRAPHIC PENCIL, following same directions as for Section V. MARK ALL SIX QUALIFICATIONS.

	FOR RATING OFFICER										FOR INDORSING OFFICER									
The degree to which he is able to meet situations without bias and without emotional upset.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
The degree to which he is able and willing to work with other officers and enlisted men.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
The degree to which he is able to act on his own responsibility in absence of orders.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
The degree to which he is able to discriminate & evaluate facts to arrive at logical conclusions.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
The degree to which his appearance and behavior cause people to react favorably.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
The degree to which he is able to carry out orders with consistency & firmness to achieve objectives.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10

Section VIII. OVER-ALL RELATIVE RANK

FOR RATER ONLY

The number of officers in this grade rated by me at this time is_____.	If these officers were arranged in order, considering over-all future usefulness to the Army, from highest (No.1) to poorest, this officer would be No._____ of the total group rated.

Section IX. AUTHENTICATION

Use typewriter (except for signatures) or ink.

I certify that I have read the current AR 600-185 and that all ratings are made in accordance with instructions contained therein, and that to the best of my knowledge and belief all entries contained hereon are true and impartial.

SIGNATURE OF RATING OFFICER	SIGNATURE OF INDORSING OFFICER
NAME, GRADE, AND ORGANIZATION OR UNIT	NAME, GRADE, AND ORGANIZATION OR UNIT
OFFICIAL STATUS OF RATED OFFICER WITH RESPECT TO RATING OFFICER	OFFICIAL STATUS OF RATED OFFICER WITH RESPECT TO INDORSING OFFICER

APPENDIX F

AIR FORCE RATING WORK BOOKLET

(SAMPLE SHEET)

40. SUBORDINATING PERSONAL INTERESTS (See also V-49)

☐	1	2	3	4	5
Unknown	Allowed personal interest and welfare to interfere seriously with performance of duty.	Made excessive use of the privileges of his rank to increase personal comfort or avoid inconvenience.	Accepted minor personal inconvenience in order to perform his duties.	Subordinated personal desires to the performance of duty.	Disregarded personal welfare and interests completely when necessary for the successful performance of duty.

41. COOPERATING WITH ASSOCIATES (See also I-11)

☐	1	2	3	4	5
Unknown	Refused to help or cooperate with an associate.	Cooperated only grudgingly with an associate.	Gave some assistance to an associate when requested to do so.	Cooperated willingly with an associate when called upon.	Voluntarily assisted an associate when help was needed.

42. SHOWING LOYALTY

☐	1	2	3	4	5
Unknown	Openly criticized a superior or associate.	Allowed a superior or associate to be criticized without coming to his defense.	Supported the policies or actions of a superior or associate.	Gave credit to a superior for accomplishments of the organization.	Shared any criticism or blame directed at a superior for the mistakes or shortcomings of his organization.

43. TAKING RESPONSIBILITY FOR SUBORDINATES (See also II-26)

☐	1	2	3	4	5
Unknown	Failed to take any responsibility for the conduct and fair treatment of a subordinate.	Took very little responsibility for the conduct and fair treatment of a subordinate.	Assumed responsibility for the conduct and fair treatment of a subordinate.	Assumed substantial responsibility for the conduct and fair treatment of a subordinate.	Assumed full responsibility for the conduct and fair treatment of a subordinate.

V. PERSONAL HABITS AND ADJUSTMENT

44. ATTENDING TO DUTY

☐	1	2	3	4	5
Unknown	Required constant prodding and close supervision to keep his attention on his assigned duties.	Worked just hard enough to get by.	Showed satisfactory industry.	Worked hard and willingly to achieve objectives.	Did extra work voluntarily in order to achieve objectives.

45. ATTENDING TO DETAILS

☐	1	2	3	4	5
Unknown	Neglected an important detail of his job with serious results.	Neglected routine details of his work, slowing up operations.	Handled satisfactorily the important details of his job.	Gave careful attention to the important details of his job.	Made sure that all details of his job were completely taken care of.

46. REPORTING FOR APPOINTMENTS

☐	1	2	3	4	5
Unknown	Failed to keep an appointment causing serious inconvenience to others.	Was late for an appointment without notifying others.	Notified others when he expected to be late for an appointment.	Made a special effort to be prompt in reporting for an appointment.	Kept an appointment in spite of substantial personal inconvenience.

47. MEETING COMMITMENTS (See also I-3)

☐	1	2	3	4	5
Unknown	Failed to meet a commitment hindering the work of others.	Was late in meeting a commitment causing inconvenience to others.	Met a commitment promptly and fully.	Met a commitment fully in spite of difficulties beyond his control.	Completed an assignment ahead of time and more fully than required.

48. IMPROVING EFFECTIVENESS (See also VI-55)

☐	1	2	3	4	5
Unknown	Refused an opportunity offered him to improve his effectiveness.	Ignored an opportunity to improve his proficiency or potentiality.	Accepted an opportunity to improve his proficiency or potentiality.	Was alert to an opportunity to improve his effectiveness.	Sought out actively an opportunity to improve his proficiency and potentiality.

49. BEING FAIR AND SCRUPULOUS (See also IV-40)

☐	1	2	3	4	5
Unknown	Made an inaccurate or evasive statement or indulged in a "sharp" practice.	Used an indirect and questionable method.	Was satisfactorily accurate, frank, or fair in a statement or practice.	Used a direct and forthright method.	Was scrupulously accurate, frank, or fair in e statement or practice.

30. MAKING USE OF EXPERIENCE (See also VI-56)

☐	1	2	3	4	5
Unknown	Made no application of his own or others' experience in similar work resulting in unnecessary delays and errors.	Failed to make adequate application of his own or others' experience resulting in a loss of effectiveness.	Made some application of his own or others' experience to assist him on a problem.	Made good use of his own or others' experience to work out a problem.	Applied his own and others' experience to the fullest extent in the solution of a problem.

31. LONG-RANGE PLANNING

☐	1	2	3	4	5
Unknown	Failed to prepare plans to guide the organization's work.	Prepared only inadequate plans based on a sketchy understanding of the objectives.	Prepared plans based on only the immediate objectives of the organization.	Prepared careful plans based on a full understanding of all the objectives.	Developed very effective long-range plans based on a thorough analysis of all objectives and of means of reaching them.

32. TAKING PROMPT ACTION

☐	1	2	3	4	5
Unknown	Refused to act when a decision was urgently needed.	Hesitated or put off making a needed decision although possessed of all essential facts.	Took necessary action without serious delay.	Took prompt action to meet an established need.	Took prompt action to meet a legitimate need even though he might have delayed.

33. SUSPENDING JUDGMENT

☐	1	2	3	4	5
Unknown	Refused to wait for or examine essential additional facts before deciding.	Failed to request or examine needed additional facts before deciding.	Decided only after he had been supplied with some additional information.	Withheld his decision on an important matter until the main additional facts needed could be examined.	Insisted in spite of strong pressure on obtaining needed additional facts before making an important decision.

34. MAKING CORRECT DECISIONS

☐	1	2	3	4	5
Unknown	Made a bad decision which appeared quite contrary to the known facts.	Made a serious error in judging the relative importance of several factors in deciding.	Made an adequate decision based on a reasonable interpretation of the facts.	Made a good decision showing sound evaluation of all the factors involved.	Made an excellent decision which exactly fitted all the factors involved.

35. PERSISTING IN EFFORTS

☐	1	2	3	4	5
Unknown	Gave up as soon as he met opposition or difficulties in pursuing an objective.	Gave up after a few attempts had failed.	Kept on working toward an objective of the organization despite minor difficulties or opposition.	Renewed his efforts after a major set-back in pursuing an objective of the organization.	Persisted in his efforts to achieve an objective of the organization despite repeated set-backs or severe opposition.

36. MAKING FORCEFUL EFFORTS

☐	1	2	3	4	5
Unknown	Made very feeble and ineffectual efforts to achieve an objective.	Made only weak and partially effective efforts to achieve an objective.	Was somewhat vigorous and forceful in his efforts to achieve an objective.	Pursued an objective of the organization with energy and force.	Made very vigorous and forceful efforts to achieve an objective.

37. ABSORBING MATERIALS

☐	1	2	3	4	5
Unknown	Delayed operations because of his poor memory or the slowness with which he absorbed facts.	Made a mistake due to slow learning or faulty memory for facts.	Learned and remembered well enough to do a satisfactory job.	Achieved good results because he learned or recalled facts better than others.	Achieved exceptional results because of his unusual ability to learn and remember pertinent information.

IV. ACCEPTANCE OF TEAM PRINCIPLE AND ORGANIZATIONAL DISCIPLINE

38. COMPLYING WITH ORDERS AND DIRECTIVES

☐	1	2	3	4	5
Unknown	Disobeyed an order or directive.	Delayed or tried to avoid compliance with an order or directive.	Complied satisfactorily with an order or directive without undue delay.	Carried out an order or followed a directive promptly.	Carried out promptly and effectively the spirit and intent of an order or directive.

39. ACCEPTING ORGANIZATIONAL PROCEDURE

☐	1	2	3	4	5
Unknown	Violated an organizational procedure or regulation.	Was openly critical of and followed only partially an organizational procedure.	Conformed without open criticism to an organizational procedure with which he had previously expressed disagreement.	Accepted a procedure fully because he understood the necessity for organizational control.	Accepted and promoted understanding of a procedure as essential to effective organizational control.

APPENDIX G

AIR FORCE RATING SHEET

Last Name	First Name	Initial	Grade	Serial No.

INSTRUCTIONS: Place an "X" in the appropriate box to denote your evaluation of the performance of this officer with respect to the critical requirements listed below. The six columns of boxes correspond to the unknown box and the five numbers on each scale in the *Check List of Critical Requirements for Officer Evaluation* on *Report of Officer Effectiveness (AF Form 77A)*.

D

I. PROFICIENCY IN HANDLING ADMINISTRATIVE DETAILS

U 1 2 3 4 5

1. Understanding instructions
2. Scheduling work
3. Getting information from records
4. Getting ideas from others
5. Checking accuracy of work
6. Writing letters and reports
7. Getting cooperation
8. Presenting finished work
9. Keeping records
10. Keeping others informed
11. Rendering effectiveness reports

X 1 2 3 4 5 X

(DO NOT USE THIS SPACE) →

II. PROFICIENCY IN SUPERVISING PERSONNEL

U 1 2 3 4 5

12. Matching personnel and jobs
13. Delegating authority
14. Giving orders and instructions
15. Insuring comprehension
16. Giving reasons and explanations
17. Supporting authorized actions
18. Encouraging ideas
19. Developing teamwork
20. Setting a good example
21. Assisting subordinates in their work
22. Evaluating subordinates' work
23. Looking out for subordinates' welfare
24. Maintaining relations with subordinates

X 1 2 3 4 5 X

(DO NOT USE THIS SPACE) →

III. PROFICIENCY IN PLANNING AND DIRECTING ACTION

U 1 2 3 4 5

25. Taking responsibility
26. Solving problems
27. Making use of experience
28. Long-range planning
29. Taking prompt action
30. Suspending judgment
31. Making correct decisions
32. Making forceful efforts
33. Absorbing materials

X 1 2 3 4 5 X

(DO NOT USE THIS SPACE) →

IV. ACCEPTANCE OF ORGANIZATIONAL RESPONSIBILITY

U 1 2 3 4 5

34. Complying with orders and directives
35. Accepting organizational procedure
36. Subordinating personal interests
37. Cooperating with associates
38. Showing loyalty
39. Taking responsibility for subordinates

X 1 2 3 4 5 X

(DO NOT USE THIS SPACE) →

V. ACCEPTANCE OF PERSONAL RESPONSIBILITY

U 1 2 3 4 5

40. Attending to duty
41. Attending to details
42. Reporting for appointments
43. Meeting commitments
44. Being fair and scrupulous
45. Maintaining military appearance
46. Adapting to associates
47. Adapting to the job
48. Conforming to civil standards

X 1 2 3 4 5 X

(DO NOT USE THIS SPACE) →

VI. PROFICIENCY IN DUTY MILITARY OCCUPATIONAL SPECIALTY

(Describe the specific duty performed and give the SSN/ of the duty)

U 1 2 3 4 5

49. Possessing fundamental training
50. Improving effectiveness
51. Keeping well-informed
52. Applying training and information
53. Showing ingenuity in specialty
54. Handling related assignments

X 1 2 3 4 5 X

(DO NOT USE THIS SPACE) →

TOTAL SCORE	(DO NOT USE THIS SPACE)

☆U S GOVERNMENT PRINTING OFFICE 1949—O 841770

APPENDIX H

MARINE CORPS RATING FORM

NAVMC 652-PD

REPORT ON FITNESS OF OFFICERS OF THE UNITED STATES MARINE CORPS

(To be submitted in accordance with Art. 137, U. S. Navy Regulations, 1920, and Art. 10-22, Marine Corps Manual)

_____, _____ U. S. M. C.
(Name—Surname first) (Rank)

Ship or station _____

Period covered _____ months, from _____ to _____

To be answered by officer reported on:

1. Regular duties _____

2. Additional duties _____

3. Wife's address _____

4. Name, relationship, and address of person other than wife to be notified in case of emergency _____

_____

_____, _____ U. S. M. C.
(Signature) (Rank)

To be answered by reporting officer:

5. Reporting officer _____, _____ U. S. _____
(Name) (Rank)

6. *Method of rating.*—When rating this officer, consider carefully and keep in mind the following definitions, taking into consideration his length of service, the opportunities afforded him which might have a bearing on his performance of duty, his personal characteristics, and professional qualifications:

   UNSATISFACTORY.—Inefficient; below minimum standard.
   FAIR.—Satisfactory; passably efficient; up to minimum standard.
   GOOD.—Average qualifications; efficient, but to a less degree than "Very good."
   VERY GOOD.—Above average; efficient; well qualified.
   EXCELLENT.—Highly efficient; qualified to a high degree.
   OUTSTANDING.—Superior; exceptionally efficient; qualified to a preeminent degree.
   NOT OBSERVED.—To be used in all cases where the reporting officer has had insufficient opportunity to observe the officer reported on during the period covered by this report to permit a rating as to performance of a particular duty, personal characteristics, or professional qualifications.

7. Before making out this report, decide in your own mind on an actual officer in the grade of the officer now being reported on who, in your opinion, based on personal knowledge, is the outstanding officer of his rank in the Marine Corps; or
   Decide in your own mind the character attributes and professional qualifications which the ideal officer in the grade of the officer now being reported on should possess.

8. Considering the officer reported on in comparison with your ideal (7), and having in mind the instructions under (6) "Method of Rating", indicate your estimate of him by marking "X" in the appropriate space below.	Not observed	Unsatisfactory	Fair	Good	Very good	Excellent	Outstanding
Performance of duty (based on fact):							
(*a*) Regular duties							
(*b*) Additional duties							
(*c*) Administrative duties							
(*d*) Executive duties							
(*e*) Handling officers							
(*f*) Handling enlisted men							
(*g*) Training troops							
(*h*) Tactical handling of troops (unit appropriate to officer's grade)							

16—18250-2

9. To what degree has he exhibited the following qualifications? Consider him in comparison with your ideal (7), and indicate your estimate by marking "X" in the appropriate space below.	Not observed	Unsatisfactory	Fair	Good	Very good	Excellent	Outstanding
(*a*) **Physical fitness** (physical stamina; endurance under hardship, adversity, or discouragement)							
(*b*) **Military bearing and neatness** (dignity of demeanor; neat and smart appearance)							
(*c*) **Attention to duty** (industry; the trait of working thoroughly and conscientiously)							
(*d*) **Cooperation** (the faculty of working in harmony with others, military or civilian)							
(*e*) **Initiative** (the trait of taking necessary or appropriate action on own responsibility)							
(*f*) **Intelligence** (the ability to grasp readily situations and instructions)							
(*g*) **Judgment and common sense** (the ability to think clearly and arrive at logical conclusions)							
(*h*) **Presence of mind** (the ability to think and act promptly and effectively in an unexpected emergency or under great strain)							
(*i*) **Force** (the faculty of carrying out with energy and resolution that which is believed to be reasonable, right, or duty)							
(*j*) **Leadership** (the capacity to direct, control, and influence others and still maintain high morale)							
(*k*) **Loyalty** (the quality of rendering faithful and willing service, and unswerving allegiance under any and all circumstances)							

10. Has he any characteristics—temperamental, moral, physical, etc.—which adversely affect his efficiency? .....
If yes, briefly describe them .....

11. During the period covered by this report, has the work of this officer been reported on either in a commendatory way, or adversely? If so, indicate subject matter and date .....

12. During the period covered by this report was he the subject of any disciplinary action that should be included on his record? ..... If yes, and if not previously reported to Headquarters, attach separate statement of nature and attendant circumstances.

13. In case any unfavorable entries have been made by you on this or on a previous report, were the deficiencies noted brought to the attention of the officer concerned? ..... If yes, what improvement, if any, has been noted? .....
If no improvement was noted, what period of time has elapsed since the deficiencies were brought to his notice? .....

14. Considering the possible requirements of the service in war, indicate your attitude toward having this officer under your command. Would you—
(*a*) Particularly desire to have him? ..... (*c*) Be willing to have him? .....
(*b*) Be glad to have him? ..... (*d*) Prefer not to have him? .....
If (*d*), explain briefly.....

15. (To be answered only when reporting on officers serving under revocable commissions.) Do you recommend retention in the service after expiration of revocable period of commission? .....
(Yes or no; if negative give reasons)

16. REMARKS: (To be used for additional pertinent information or comment, if any, not covered elsewhere in this report)

17. Indicate your estimate of this officer's "General Value to the Service", using the ratings specified in (6) .....

18. Having in mind the special fitness of this officer and the efficiency of the naval service, I certify that to the best of my knowledge and belief all entries made hereon are true and without prejudice or partiality.

_____, _____ U. S. _____
(Signature) (Rank)

_____ _____
(Date) (Duty)

U.S. GOVERNMENT PRINTING OFFICE 16—18250

APPENDIX I

GLOSSARY OF TERMS

APPENDIX I

GLOSSARY OF TERMS

Behavior -- Any or all of the overt actions of an individual including manifestations of emotions. In the field of efficiency rating it refers particularly to those actions or manifestations which are significant to the job being performed by the ratee.

Cognition-- A concept, a mental picture. Also the act of forming a mental picture.

Continuum-- The relation which exists when several entities are actually varying degrees of the same quality of thing.

Correlation-- The degree with which two phenomena are related or associated. Correlation implies association only, not causality.

Correlation, Coefficient of-- The mathematical expression of the degree of association. A coefficient of 1.00 signifies perfect association for, knowing the value of one phenomenon, the personnel manager may predict the value of the other exactly. Coefficients of less than 1.00 indicate that prediction will be less than perfect until, at a coefficient of 0.0,

APPENDIX I

GLOSSARY OF TERMS

Behavior -- Any or all of the overt motions of an individual including manifestations of emotions. In the field of efficiency rating it refers particularly to those motions or manifestations which are significant to the job being performed by the rater.

Cognition-- Knowing, a mental picture. Also the act of forming a mental picture.

Continuum-- The relation which exists between all things are actually varying degrees of the same reality or thing.

Correlation-- The degree of relation between phenomena or variables. Correlation implies association only, not causality.

Correlation, coefficient of-- The mathematical expression of the degree of association or interdependence of two or more factors. Correlation, in the sense of one phenomenon, disappearance and or by [illegible] of [illegible] expect . [illegible] of [illegible]

there is no relation and prediction will be no better than blind guessing. Negative coefficients have predictive value comparable to that of positive coefficients of the same numerical magnitude. However, with a negative coefficient, the maximum display of one phenomenon is associated with the minimum of the other and prediction must be made on that basis.

Distribution -- Any group of scores obtained on the same rating or testing device and gathered together for consideration.

Efficiency Rating -- This is one name for an assessment of that portion of an individual's value on a job which is made up of intangible qualities not measurable by more objective means. The term is also applied to the blank paper form used to systematize and record this assessment.

Efficiency Report -- The U. S. Army efficiency rating device for officers. The term "report" carries the connotation that only a <u>description</u> of the ratee is being made with the derivation of a <u>rating</u> being postponed until the report reaches headquarters.

Fitness Report -- The U. S. Navy equivalent of the Army Efficiency Report.

there is by reflection and pyrolysis still no better than field measure. The [illegible] coefficients have relative values comparable to the [illegible] tive coefficients of the [illegible] magnetic. However, with a magnetic coefficient, the [illegible] dependency of the phenomena is associated with the [illegible] of the other [illegible] to [illegible] on [illegible] model.

Dielectrics -- Any atom of square contained [illegible] ration or testing device [illegible] to other for comparison.

Efficiency unit -- This is one [illegible] of the particular of an [illegible] value on a [illegible] which is one of [illegible] the measurable or more objective means. The [illegible] applied to the [illegible] to [illegible] and [illegible].

Efficiency report -- [illegible] efficiency [illegible] for efficiency. The term "[illegible]" [illegible] more [illegible] only a [illegible] of [illegible] data [illegible] a [illegible] of [illegible] correspondence [illegible].

Efficiency [illegible] -- [illegible]

Mean -- The arithmetical average of a group of numbers.

Median -- The point or number in a distribution which divides the distribution into two equal parts.

Merit Rating -- The Industrial psychologist's equivalent of the Efficiency Rating.

N -- The symbol for the number of subjects participating in an experiment.

Perception -- That which takes place when an individual becomes aware of the occurrence of a phenomenon or of the existence of an entity.

Percentile Score -- A score presented in such form that it shows the percentage of the entire distribution which lies below that particular score. Thus a 75th percentile score is higher than 75% of the scores of that distribution.

Ratee -- The individual who is the subject of the efficiency rating or report.

Rater -- The person who performs the assessment and makes out the rating form.

Raw Score -- The actual score attained on a rating.

Mean -- The arithmetical average of a group of measures.

Median -- The point or measure in a distribution which divides the distribution into two equal parts.

Merit Rating -- The industrial psychologist's evaluation of the efficiency rating.

N -- The symbol for the number of subjects participating in an experiment.

Perception -- That which takes place when an individual becomes aware of the existence of a stimulation or of the existence of an entity.

Personality -- A term pertaining to the total man. It embraces the personal and social adjustments, such like color, other particular traits. There is much personality among the [illegible] 78 of the society of body identifications.

Psyche -- [illegible]

[illegible] -- [illegible]

[illegible] -- [illegible]

Reliability -- That quality of a measuring instrument which causes the device to produce the same result on successive measurements of the same entity.

Reporting Senior -- The Navy's term for the Rater.

Service Rating -- The name applied to efficiency rating in the field of public personnel administration.

Service Report -- The name applied to the efficiency report in the field of public personnel administration.

Skewed Distribution -- A distribution which has departed from the normal bell-shaped curve and has the major portion of its scores concentrated near one end of the scale.

Split-Halves Reliability -- A kind of reliability determined by dividing the test or rating device into two equivalent parts and correlating the scores on the two parts. The resulting coefficient is known as the split-halves reliability coefficient.

Standard Deviation -- A statistical concept; a measure of the variability of the distribution. It is determined by taking the square root of the average of the deviations (from the mean) of all the scores in the distribution.

Reliability -- That quality of a measuring instrument which causes the device to produce the same results on successive measurements of the same entity.

Reporting order -- The Navy's term for the chart.

Service medal -- The name applied to efficiency ratings in the field of enlisted personnel evaluation.

Service record -- The name applied to the efficiency report in the field of enlisted personnel evaluation.

Shared Identification -- A distinction which is separated from the normal well-shaped curve and [illegible] a major portion of its general concentration near one end of the scale.

Split-half reliability -- A kind of reliability determined by dividing the items or rating series into two equivalent parts and correlating the [illegible] results. [illegible] coefficient is known as a equivalent-reliability coefficient.

[illegible] -- [illegible]

Standard Score -- The standard score equivalent of a raw score is the deviation of that raw score from the mean of the scores, divided by the standard deviation of the distribution. The standard score is a ratio and as such is comparable to standard scores in other distributions. Therein lies its usefulness.

Statistical Significance -- A term applied to describe a difference between two measures. It refers to a convention observed by statisticians whereby, if the difference is three times as great as the standard deviation of that difference, the chance that on repeated measurements the difference will be found to occur in the opposite direction is so small as to be negligible. Under such conditions, the difference is said to be "statistically significant".

Subject -- The general term for the individual being subjected to study in an experiment.

Trait -- A consistency of behavior toward a given stimulus.

Validity -- The ability of a measuring device to measure that which it purports to measure.

Standard score -- The standard score equivalent of a raw score is the deviation of that raw score from the mean of the series, divided by the standard deviation of the distribution. The standard score is a ratio and as such is comparable to another score of some distribution, regardless of the units.

Statistical significance -- A term applied to describe a difference between two measures. It refers to a condition observed by statisticians somebody if the difference in score from a given to the standard deviation of that difference. The course for an experiment measurement the difference will be found to occur in a specific direction. It is said to exist regardless. Thus the "statistically significant" difference by well.

Subject -- The general term for the individual subjected to physical or mental tests.

Trait -- [illegible]

Utility -- [illegible]

Validity, Face -- Validity which is awarded to a device by reason of the fact that it appears (on consideration of its makeup) to measure that which it purports to measure.

z Score -- A synonym for Standard Score.

Validity, Face -- Validity which is awarded to a device b. reason of the fact that it appears (on consideration of its making) to measure that which it purports to measure.

z Score -- A synonym for standard score.

U.S.N.A.P.
89

OCT 18 3749
JE 22 56 13078
MY 16 57 7152
MY 12 10050
DE 9 59 10814
MR 13 63 13518

Thesis 12839
F54 Foley
The adequacy of
officer fitness reports

Thes
F54

MY 16 57 7152
MY 12 59 10050
DE 9 59 10814

thesF54
The adequacy of officer fitness reports.
3 2768 001 96825 8
DUDLEY KNOX LIBRARY